

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/34809 A2

- (51) International Patent Classification⁷: **C12N 15/31**, C07K 14/31, 16/12, C12Q 1/68, G01N 33/53, A61K 39/085, 48/00
- (74) Agents: **LEVY, David, J.**; Glaxo Wellcome Inc., Five Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709-3398 et al. (US).
- (21) International Application Number: PCT/US00/30782
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date:
9 November 2000 (09.11.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (30) Priority Data:
60/164,258 9 November 1999 (09.11.1999) US
- (71) Applicant (*for all designated States except US*): **GLAXO GROUP LIMITED** [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB).
- Published:**
— Without international search report and to be republished upon receipt of that report.
- (72) Inventor; and
- (75) Inventor/Applicant (*for US only*): **KIMMERLY, William, John** [US/US]; 242 Via Villena, Encinitas, CA 92024 (US).
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/34809 A2

(54) Title: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

(57) Abstract: S epidermidis polypeptides and DNA (RNA) encoding such polypeptides and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing such polypeptides and DNA (RNA) for the treatment of infection, particularly infections arising from S epidermidis. Antagonists against the function of such polypeptides and their use as therapeutics to treat infection are also disclosed. Also disclosed are diagnostic assays for detecting diseases related to the presence of S epidermidis nucleic acid sequences and the polypeptides in a host. Also disclosed are diagnostic assays for detecting polynucleotides and polypeptides related to S epidermidis.

STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

Field of the Invention

The present invention provides nucleic acids, and peptides, polypeptides and proteins encoded by the nucleic acids, isolated from *Staphylococcus*
5 *epidermidis*.

Background of the Invention

Staphylococcus epidermidis is a gram-positive bacteria present in the normal flora of humans, and is typically present on the skin. It is catalase positive
10 and grows aerobically. It is implicated in various human conditions and diseases, including subacute bacterial endocarditis (Baddour LM et al., Production of experimental endocarditis by coagulase-negative staphylococci: variability in species virulence, J. Infect. Dis. 150: 721-727, 1984; Karchmer AW, Archer GL, Dismukes WE, *Staphylococcus epidermidis* causing prosthetic valve endocarditis: microbiologic and clinical observations as guides to therapy, Ann Intern Med.
15 1983;98:447-455.) and septicemia (Christensen GD et al., Nosocomial septicemia due to multiply antibiotic-resistant *Staphylococcus epidermidis*, Ann. Intern. Med. 96: 1-10, 1982). *S. epidermidis* is estimated to be responsible for about 12% of all hospital patient infections. Because of the organism's peculiar ability to colonize
20 polymer and metallic surfaces, there is a correlation of infection with the insertion of intravenous lines or catheters or implantation of prosthetic devices. Treatment can be difficult since different isolates of *S. epidermidis* show a broad spectrum of antibiotic resistance. The organism also produces a polysaccharide biofilm which helps to protect the bacteria from the human immune system (Tojo M et al.,
25 Isolation and characterization of a capsular polysaccharide adhesin from *Staphylococcus epidermidis*, J. Infect. Dis. 157: 713-722, 1988).

The present invention advantageously provides isolated nucleic acids and their encoded peptides, polypeptides and proteins from the genome of *S. epidermidis*, as well as the genomic map of *S. epidermidis*. Thus, the present
30 invention fulfills a a widely-felt need for *S. epidermidis* diagnostics, antigens, and

products useful in procedures for preparing antibodies and for identifying compounds effective against *S. epidermidis* infection. Selected nucleic acids and/or polypeptides of the present invention can be advantageously utilized as targets in screenings assays for antibiotics, as diagnostics of infections, and as means to identify *S. epidermidis* in any given sample and distinguish it from other bacteria.

SUMMARY OF THE INVENTION

The present invention provides an isolated polynucleotide comprising a member selected from the group consisting of:

- (a) a polynucleotide encoding a polypeptide having at least a 70% identity to a polypeptide set forth in the Sequence Listing;
- (b) a polynucleotide which is complementary to the polynucleotide of (a); and
- (c) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a) or (b). The present invention further provides polypeptides encoded by these polynucleotides and methods of using the polynucleotides and polypeptides.

DETAILED DESCRIPTION OF THE INVENTION

GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the Examples. The explanations are provided as a convenience and are not limitative of the invention.

BINDING MOLECULE refers to a molecule or ion which binds or interacts specifically with polypeptides or polynucleotides of the present invention, including, for example enzyme substrates, cell membrane components and classical receptors. Binding between polypeptides (or polynucleotides) of the invention and such molecules may be exclusive to polypeptides of the invention, which is preferred, or it may be highly specific for polypeptides of the invention, which is also preferred, or it may be highly specific to a group of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes a polypeptide of the invention. Binding molecules also include antibodies and antibody-derived reagents that bind specifically to polypeptides of the invention.

GENETIC ELEMENT generally means a polynucleotide comprising a region that encodes a polypeptide or a polynucleotide region that regulates replication, transcription or translation or other processes important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression. Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within plasmids. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

HOST CELL is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

IDENTITY, as known in the art, is the relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Identity can be readily calculated (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or two polypeptide sequences, the term is well known to skilled artisans (*Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988)). Methods commonly employed to determine identity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48:1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity are codified in computer programs. Preferred computer

program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403 (1990)).

5 ISOLATED means separated "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living organism in its natural state is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials
10 of its natural state is "isolated", as the term is employed herein. As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in
15 whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions
20 or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or
25 DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more
30 typically, double-stranded, or triple-stranded, or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a
35 region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term polynucleotide includes DNAs or

5 RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA
10 characteristic of viruses and cells, including simple and complex cells, *inter alia*. The term polynucleotide also embraces short polynucleotides often referred to as oligonucleotide(s). "Polynucleotide" and "nucleic acid" are often used interchangeably herein..

POLYPEPTIDES, as used herein, includes all polypeptides as described
15 below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, unless otherwise indicated, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to
20 longer chains, which generally are referred to in the art as proteins, of which there are many types. It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many amino acids, including the terminal amino acids, may be modified in a given polypeptide, either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well described in basic texts and in more detailed monographs, as well as in a
30 voluminous research literature, and they are well known to those of skill in the art. Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization,
35 disulfide bond formation, demethylation, formation of covalent cross-links, formation

of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance

10 *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS* B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be generally as a result of

20 posttranslational events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-

25 chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli* or other cells, prior to proteolytic processing,

30 almost invariably will be N-formylmethionine. During post-translational modification of the peptide, a methionine residue at the NH₂-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of the protein of the invention. The modifications that occur in a polypeptide often will be a function of how it is

35 made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host

cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as, for example, *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cell often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation, *inter alia*. Similar considerations apply to other modifications. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized recombinantly by expressing a polynucleotide in a host cell.

VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that differ from a reference polynucleotide or polypeptide, respectively. Variants in this sense are described below and elsewhere in the present disclosure in greater detail. (1) A polynucleotide that differs in nucleotide sequence from another, reference polynucleotide. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. As noted below, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. (2) A polypeptide that differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Techniques are available to evaluate temporal gene expression in bacteria, particularly as it applies to viability under laboratory and host infection conditions. A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of expression of a sequence by one of these methods yields additional information about its function and permits the selection of such sequence for further development as a screening target. Briefly, these approaches include:

1) Signature Tagged Mutagenesis (STM)

This technique is described by Hensel *et al.*, *Science* 269: 400-403(1995), the contents of which is incorporated by reference for background purposes. Signature tagged mutagenesis identifies genes necessary for the establishment/maintenance of infection in a given infection model.

The basis of the technique is the random mutagenesis of target organism by various means (e.g., transposons) such that unique DNA sequence tags are inserted in close proximity to the site of mutation. The tags from a mixed population of bacterial mutants and bacteria recovered from an infected host are detected by amplification, radiolabeling and hybridization analysis. Mutants attenuated in virulence are revealed by absence of the tag from the pool of bacteria recovered from infected hosts.

In *Streptococcus pneumoniae*, because the transposon system is less well developed, a more efficient way of creating the tagged mutants is to use the insertion-duplication mutagenesis technique as described by Morrison *et al.*, *J. Bacteriol.* 159:870 (1984) the contents of which is incorporated by reference for background purposes.

2) In Vivo Expression Technology (IVET)

This technique is described by Camilli *et al.*, *Proc. Nat'l. Acad. Sci. USA.* 91:2634-2638 (1994), the contents of which is incorporated by reference for background purposes. IVET identifies genes up-regulated during infection when compared to laboratory cultivation, implying an important role in infection. Sequences identified by this technique are implied to have a significant role in infection establishment/maintenance.

In this technique random chromosomal fragments of target organism are cloned upstream of a promoter-less reporter gene in a plasmid vector. The pool is introduced into a host and at various times after infection bacteria may be recovered and assessed for the presence of reporter gene expression. The

chromosomal fragment carried upstream of an expressed reporter gene should carry a promoter or portion of a gene normally upregulated during infection. Sequencing upstream of the reporter gene allows identification of the up regulated gene.

5 3) Differential display

 This technique is described by Chuang *et al.*, *J. Bacteriol.* 175:2026-2036 (1993), the contents of which is incorporated by reference for background purposes. This method identifies those genes which are expressed in an organism by identifying mRNA present using randomly-primed RT-PCR. By
10 comparing pre-infection and post infection profiles, genes up and down regulated during infection can be identified and the RT-PCR product sequenced and matched to library sequences.

 4) Generation of conditional lethal mutants by transposon mutagenesis.

 This technique, described by de Lorenzo, V. *et al.*, *Gene* 123:17-24
15 (1993); Neuwald, A. F. *et al.*, *Gene* 125: 69-73(1993); and Takiff, H. E. *et al.*, *J. Bacteriol.* 174:1544-1553(1992), the contents of which is incorporated by reference for background purposes, identifies genes whose expression are essential for cell viability.

 In this technique transposons carrying controllable promoters, which
20 provide transcription outward from the transposon in one or both directions, are generated. Random insertion of these transposons into target organisms and subsequent isolation of insertion mutants in the presence of inducer of promoter activity ensures that insertions which separate promoter from coding region of a gene whose expression is essential for cell viability will be recovered.
25 Subsequent replica plating in the absence of inducer identifies such insertions, since they fail to survive. Sequencing of the flanking regions of the transposon allows identification of site of insertion and identification of the gene disrupted. Close monitoring of the changes in cellular processes/morphology during growth in the absence of inducer yields information on likely function of the gene. Such
30 monitoring could include flow cytometry (cell division, lysis, redox potential, DNA replication), incorporation of radiochemically labeled precursors into DNA, RNA, protein, lipid, peptidoglycan, monitoring reporter enzyme gene fusions which respond to known cellular stresses.

 5) Generation of conditional lethal mutants by chemical mutagenesis.

35 This technique is described by Beckwith, J., *Methods in Enzymology* 204:

3-18(1991), the contents of which are incorporated herein by reference for background purposes. In this technique random chemical mutagenesis of target organism, growth at temperature other than physiological temperature (permissive temperature) and subsequent replica plating and growth at different temperature (e.g. 42°C to identify ts, 25°C to identify cs) are used to identify those isolates which now fail to grow (conditional mutants). As above close monitoring of the changes upon growth at the non-permissive temperature yields information on the function of the mutated gene. Complementation of conditional lethal mutation by library from target organism and sequencing of complementing gene allows matching with library sequences.

Each of these techniques may have advantages or disadvantages depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind. For example, some genes might be recognised as essential for infection but in reality are only necessary for the initiation of infection and so their products would represent relatively unattractive targets for antibacterials developed to cure established and chronic infections.

6) RT-PCR

Bacterial messenger RNA, preferably that of *S. epidermidis*, is isolated from bacterial infected tissue e.g. 48 hour murine lung infections, and the amount of each mRNA species assessed by reverse transcription of the RNA sample primed with random hexanucleotides followed by PCR with gene specific primer pairs. The determination of the presence and amount of a particular mRNA species by quantification of the resultant PCR product provides information on the bacterial genes which are transcribed in the infected tissue. Analysis of gene transcription can be carried out at different times of infection to gain a detailed knowledge of gene regulation in bacterial pathogenesis allowing for a clearer understanding of which gene products represent targets for screens for novel antibacterials. Because of the gene specific nature of the PCR primers employed it should be understood that the bacterial mRNA preparation need not be free of mammalian RNA. This allows the investigator to carry out a simple and quick RNA preparation from infected tissue to obtain bacterial mRNA species which are very short lived in the bacterium (in the order of 2 minute half-lives). Optimally the bacterial mRNA is prepared from infected murine lung tissue by mechanical disruption in the presence of TRIzol (GIBCO-BRL) for very short periods of time, subsequent processing according to the manufacturers of TRIzol reagent and

DNAase treatment to remove contaminating DNA. Preferably the process is optimized by finding those conditions which give a maximum amount of bacterial 16S ribosomal RNA, preferably that of *S. epidermidis*, as detected by probing Northern blots with a suitably labeled sequence specific oligonucleotide probe.

- 5 Typically a 5' dye labelled primer is used in each PCR primer pair in a PCR reaction which is terminated optimally between 8 and 25 cycles. The PCR products are separated on 6% polyacrylamide gels with detection and quantification using GeneScanner (manufactured by ABI).

- 10 Use of the of these technologies when applied to the sequences of the present invention enables identification of bacterial proteins expressed during infection, inhibitors of which would have utility in anti-bacterial therapy.

Polynucleotides

- The present invention relates to novel polynucleotides and novel
15 polypeptides of *S. epidermidis*, among other things, as described below. The invention particularly relates to the nucleotide sequences set forth in the Sequence Listing SEQ ID NOs: 1-3334, typically as odd numbered ID numbers, and the corresponding deduced amino acid sequences also set forth in the Sequence Listing SEQ ID NOs:1-3334, typically as even numbered ID numbers.
20 SEQ ID NOs 1-3334 refer to open reading frames (ORFs). The invention also relates to consensus polynucleotide sequences from which the ORFs were extracted. These genomic sequences include the ORFs, intergenic regions and ribosomal RNA genes. Such genomic polynucleotides are set forth as SEQ ID Nos 3335-4464. It will be noted that minor errors in sequencing can occur which
25 do not depart from the spirit of the invention; *S. epidermidis* polynucleotides and polypeptides having any corrected sequences are thus encompassed by this invention.

- Using the information provided herein and known, standard methods, such as those for cloning and sequencing and those for synthesizing polynucleotides
30 and polypeptides (see, e.g., Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)), one can generate numerous unique fragments, both longer and shorter than the polynucleotides and polypeptides set forth in the Sequence Listing, of the *S. epidermidis* genome and the *S. epidermidis* coding regions, which are
35 encompassed by the present invention. To be unique, a fragment must be of sufficient size to distinguish it from other known nucleic acid sequences, most

readily determined by comparing any selected *S. epidermidis* fragment to the nucleotide sequences in computer databases such as GenBank. Such comparative searches are standard in the art. Many unique fragments will be *S. epidermidis* – specific. Typically, a unique fragment useful as a primer or probe will be at least about 20 to about 25 nucleotides in length, depending upon the specific nucleotide content of the sequence. Additionally, fragments can be, for example, at least about 30, 40, 50, 60, 75, 80, 90, 100, 150, 200, 250, 300, 400, 500 or more nucleotides in length. The nucleic acid fragment can be single, double or triple stranded, depending upon the purpose for which it is intended.

Additionally, as discussed above and below, modifications can be made to the *S. epidermidis* polynucleotides and polypeptides that are encompassed by the present invention. For example, nucleotide substitutions can be made which do not affect the polypeptide encoded by the nucleic acid, and thus any polynucleotide which encodes the polypeptides of this invention is within the present invention. Additionally, certain amino acid substitutions (and corresponding nucleotide substitutions to encode them) can be made which are known in the art to be neutral (Robinson W.E. Jr. and Mitchell, W.m., *AIDS* 4: S141-S162 (1990). Such variations may arise naturally as allelic variations (*e.g.*, due to genetic polymorphism) or may be produced by human intervention (*e.g.*, by mutagenesis of cloned DNA sequences), such as induced point, deletion, insertion and substitution mutations. Minor changes in amino acid sequence are generally preferred, such as conservative amino acid replacements, small internal deletions or insertions, and additions or deletions at the ends of the molecules. Substitutions may be designed based on, for example, the model of Dayhoff, *et al.* (in *Atlas of Protein Sequence and Structure* 1978, Nat'l Biomed. Res. Found., Washington D.C.). These modifications can result in changes in the amino acid sequence, provide silent mutations, modify a restriction site, or provide other specific mutations. Likewise, such amino acid changes result in a different nucleic acid encoding the polypeptides and proteins. Thus, alternative polynucleotides, which are within the parameters of the present invention, are contemplated by such modifications.

Furthermore, the polynucleotide sequences set forth as SEQ ID Nos: 1-3334 in the Sequence Listing are open reading frames (ORFs), *i.e.*, coding regions of *S. epidermidis*. The polypeptide encoded by each open reading frame can be deduced, and the molecular weight of the polypeptide thus calculated using amino acid residue molecular weight values well known in the art. Any

selected coding region can be functionally linked, using standard techniques such as standard subcloning techniques, to any desired regulatory sequence, whether a *S. epidermidis* regulatory sequence or a heterologous regulatory sequence, or to a heterologous coding sequence to create a fusion protein, as further described
5 herein.

Polynucleotides of the present invention may be in the form of RNA, such as mRNA or cRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The DNA may be triple-stranded, double-stranded or single-
10 stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The coding sequence which encodes a *S. epidermidis* polypeptide of this invention may be identical to the coding sequence of a polynucleotide set forth in the sequence listing. It also may be a polynucleotide with a different sequence which, as
15 a result of the redundancy (degeneracy) of the genetic code, encodes a *S. epidermidis* polypeptide set forth in the sequence listing.

Polynucleotides of the present invention which encode a *S. epidermidis* polypeptide set forth in the sequence listing may include, but are not limited to, the coding sequence for a mature polypeptide, by itself; the coding sequence for a
20 mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of a mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the
25 transcribed, non-translated sequences that play a role in transcription (including termination signals, for example), ribosome binding, mRNA stability elements, and additional coding sequence which encode additional amino acids, such as those which provide additional functionalities. Thus, for instance, a polypeptide may be fused to a marker sequence, such as a peptide, which facilitates purification of the
30 fused polypeptide. In certain embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in the pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.
35 The HA tag may also be used to create fusion proteins and corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by

Wilson *et al.*, *Cell* 37: 767 (1984), for instance. Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated genetic elements.

In accordance with the foregoing, the term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides which include a sequence encoding a polypeptide of the present invention, particularly a polypeptide having a *S.epidermidis* amino acid sequence set forth in the Sequence Listing. The term encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The present invention further relates to variants of the herein above described polynucleotides which encode for fragments, analogs and derivatives of the polypeptide having a deduced *S. epidermidis* amino acid sequence set forth in the Sequence Listing. A variant of the polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Preferred are polynucleotides encoding a variant, analog, derivative or fragment, or a variant, analogue or derivative of a fragment, which have a *S. epidermidis* sequence as set forth in the Sequence Listing, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid(s) is substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the *S.epidermidis* polypeptides set forth in the Sequence Listing. Also especially preferred in this regard are conservative substitutions.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding a polypeptide having an amino acid sequence set forth in the Sequence Listing, and polynucleotides which are complementary to such polynucleotides. Alternatively,

most highly preferred are polynucleotides that comprise a region that is at least 80% or at least 85% identical over their entire length to a polynucleotide encoding a *S. epidermidis* polypeptide set forth in the Sequence Listing, including complementary polynucleotides. In this regard, polynucleotides at least 90%,
5 91%, 92%, 93%, 94%, 95%, or 96% identical over their entire length to the same are particularly preferred, and among these particularly preferred polypeptides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these, those with at least 98% and at least 99% are particularly highly preferred, with at least 99%
10 or 99.5% being the more preferred.

Preferred embodiments in this respect, moreover, are polynucleotides which encode polypeptides which retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA set forth in the Sequence Listing.

15 The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. Stringent conditions are typically selective conditions. As herein used, the term "stringent conditions" means hybridization will
20 occur only if there is at least 95% and preferably at least 97% identity between the sequences. For a specific sequence, stringent conditions can be determined empirically according to the nucleotide content, as is known in the art. For example, a typical example of stringent conditions is hybridization of a 48mer having 55% GC content at 42°C in 50% formamide and 750 mM NaCl followed by washing at 55°C in
25 15 mM NaCl and 0.1% SDS.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to
30 isolate cDNA and genomic clones of other genes that have a high sequence similarity to the polynucleotides of the present invention. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 20, at least 25 or at least 30 bases, and may have at least 50 bases. Particularly preferred probes will have at least 30 bases, and will have 50 bases or less, such as 30, 35, 40, 45, or
35 50 bases.

For example, the coding region of the polynucleotide of the present invention may be isolated by screening using the known DNA sequence to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine to which members of the library the probe hybridizes.

The polynucleotides and polypeptides of the present invention may be employed as reagents and materials for development of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays, *inter alia*.

The polynucleotides of the present invention that are oligonucleotides can be used in the processes herein as described, but preferably for PCR, to determine whether or not the *S. epidermidis* genes identified herein in whole or in part are present and/or transcribed in infected tissue such as blood. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The polynucleotides may encode a polypeptide which is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

The present invention additionally contemplates polynucleotides functionally encoding fusion polypeptides wherein the fusion polypeptide comprises a fragment of a *S. epidermidis* polypeptide and one or more polypeptide(s) derived from another *S. epidermidis* polypeptide or from another organism or a synthetic polyamino acid sequence. Such polynucleotides may or may not encode amino acid sequences to facilitate cleavage of the *S. epidermidis* polypeptide from the other polypeptide(s) under appropriate conditions.

In sum, a polynucleotide of the present invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences which are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Polypeptides

The present invention further relates to peptides, polypeptides and proteins (collectively referred to as "polypeptides") of *S. epidermidis*. The amino acid sequence of these polypeptides is set forth in the Sequence Listing.

The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to a polypeptide whose amino acid sequence is set forth in the Sequence Listing, means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The fragment, derivative or analog of the polypeptide of the present invention may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Among the particularly preferred embodiments of the invention in this regard are polypeptides set forth in the Sequence Listing, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments. Additionally, fusion polypeptides comprising such polypeptides, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments, in addition to a heterologous polypeptide, are contemplated by the

present invention. Such fusion polypeptides and proteins, as well as polynucleotides encoding them, can readily be made using standard techniques, including standard recombinant techniques for producing and expressing a recombinant polynucleic acid encoding a fusion protein.

5 Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl
10 residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, having the
15 amino acid sequence of any polypeptide set forth in the Sequence Listing, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the polypeptide of the present invention. Also especially preferred in this regard
20 are conservative substitutions. Most highly preferred are polypeptides having an amino acid sequence set forth in the Sequence Listing without substitutions.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The polypeptides of the present invention include any polypeptide set forth in
25 the Sequence Listing (in particular a mature polypeptide) as well as polypeptides which have at least 70% identity to a polypeptide set forth in the Sequence Listing, preferably at least 80% or 85% identity to a polypeptide set forth in the Sequence Listing, and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide set forth in the Sequence Listing and still more preferably at
30 least 95%, 96%, 97%, 98%, 99%, or 99.5% similarity (still more preferably at least 95%, 96%, 97%, 98%, 99%, or 99.5% identity) to a polypeptide set forth in the Sequence Listing and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids, such as 30, 35, 40, 45 or 50 amino acids.

35 Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide

synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

5 **Fragments**

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of the polypeptide having the amino acid sequence set forth in the Sequence Listing, and fragments of variants and derivatives of the polypeptides set forth in the Sequence Listing.

10 In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned S. epidermidis polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," *i.e.*, not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a polypeptide of the present invention comprised within a precursor polypeptide designed for expression in a host and having heterologous pre and pro-polypeptide regions fused to the amino terminus of the fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from a polypeptide of the present invention.

25 Representative examples of polypeptide fragments of the invention, include, for example, in any selected polypeptide, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, and 101-200, 201-300, or, at the COOH-terminal end, the C-terminal 20 amino acids, the C-terminal 30 amino acids, the C-terminal 40 amino acids, the C-terminal 50 amino acids, and any combination of these fragments, such as fragment from about amino acid number 1-40, 1-60, 21-60, 41-80, 61-100, and the like.

In this context "about" herein includes the particularly recited ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

35 Preferred fragments of the invention include, for example, truncation polypeptides including polypeptides having an amino acid sequence set forth in the

Sequence Listing, or of variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out above also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally. Degradation forms of the polypeptides of the invention in a host cell are also preferred.

Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of the polypeptide of the present invention Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions of the polypeptide of the present invention, and combinations of such fragments.

Preferred regions are those that mediate activities of the polypeptide of the present invention. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of the polypeptide of the present invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *S.epidermidis* or the ability to cause disease in humans. Further preferred polypeptide fragments are those that comprise or contain antigenic or immunogenic determinants in an animal, especially in a human.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspond to the preferred fragments, as discussed above.

Vectors, host cells, expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of
5 polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. Introduction of a polynucleotides into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated
10 transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
15 (1989).

Polynucleotide constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

20 Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al.,
25 *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Plasmids generally are designated herein
30 by a lower case p preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Starting plasmids disclosed herein are either commercially available, publicly available, or can be constructed from available plasmids by routine application of well known, published procedures, given the teachings herein. Many plasmids and
35 other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise *cis*-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate *trans*-acting factors either
5 are supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific.
10 Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of
15 skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements,
20 from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with
25 this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques, such as, for example, those set forth in
30 Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include, but are not
35 limited to, the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon
5 appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures, such regions will operate by controlling transcription, such as transcription factors, repressor binding sites and termination, among others.

10 Vectors for propagation and expression generally will include selectable markers and amplification regions, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

Representative examples of appropriate hosts include bacterial cells, such as
15 streptococci, staphylococci, *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

The following vectors, which are commercially available, are provided by way
20 of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia, and pBR322 (ATCC 37017). Among preferred eukaryotic vectors are pWLNEO, pSV2CAT,
25 pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for example,
30 introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("CAT") transcription unit, downstream of
35 restriction site or sites for introducing a candidate promoter fragment; *i.e.*, a fragment that may contain a promoter. As is well known, introduction into the vector of a

promoter-containing fragment at the restriction site upstream of the *cat* gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available, such as pKK232-8 and pCM7. Promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known prokaryotic promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* *lacI* and *lacZ* and promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR, PL promoters and the *trp* promoter.

Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Recombinant expression vectors will include, for example, origins of replication, a promoter preferably derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiation codon. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal in constructs for use in eukaryotic hosts. Transcription termination signal appropriately disposed at the 3' end of the transcribed region may also be included in the polynucleotide construct.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide.

These signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N- or C-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, regions also may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability or to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize or purify polypeptides. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another protein or part thereof. In drug discovery, for example, proteins have been fused with antibody Fc portions for the purpose of high-throughput screening assays to identify antagonists. See, D. Bennett *et al.*, *Journal of Molecular Recognition*, Vol. 8 52-58 (1995) and K. Johanson *et al.*, *The Journal of Biological Chemistry*, Vol. 270, No. 16, pp 9459-9471 (1995).

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents; such methods are well known to those skilled in the art.

Mammalian expression vectors may comprise expression sequences, such as an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation regions, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are useful or necessary for expression.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high

performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polynucleotide assays

5 This invention is also related to the use of the polynucleotides of the present invention to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of complementary nucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly
10 mammals, and especially humans, infected with *S. epidermidis* may be detected at the DNA level by a variety of techniques. By selecting regions of nucleic acids that vary among strains of *S. epidermidis*, preferred candidates for distinguishing a specific strain of *S. epidermidis* can be obtained. Furthermore, by selecting regions of nucleic acids that vary between *S. epidermidis* and other organisms, preferred
15 candidates for distinguishing a *S. epidermidis* from other organisms can be obtained. Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki et al., *Nature*, 324: 163-166 (1986) prior to analysis. RNA or cDNA may also be used
20 in the same ways. As an example, PCR primers complementary to the nucleic acid forming part of the polynucleotide of the present invention can be used to identify and analyze for its presence and/or expression. Using PCR, characterization of the strain of *S. epidermidis* present in a mammal, and especially a human, may be made by an analysis of the genotype of the prokaryote gene. For example, deletions and
25 insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA or alternatively, radiolabeled antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in
30 melting temperatures.

 Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or
35 another amplification method. For example, a sequencing primer can be used with double-stranded PCR product or a single-stranded template molecule generated by

a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Genetic characterization based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers et al., *Science*, 230: 1242 (1985)).

Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, e.g., restriction fragment length polymorphisms (RFLP) and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by *in situ* analysis.

Cells carrying mutations or polymorphisms in the gene of the present invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to the nucleic acid encoding the polypeptide of the present invention can be used to identify and analyze mutations. The primers may be used to amplify the gene isolated from the individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be diagnosed.

The invention provides a process for diagnosing disease, arising from infection with *S. epidermidis*, comprising determining from a sample isolated or derived from an individual an increased level of expression of a polynucleotide having the sequence of a polynucleotide set forth in the Sequence Listing. Increased expression of polynucleotide can be measured using any one of the

methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Polypeptide assays

5 The present invention also relates to diagnostic assays such as quantitative and diagnostic assays for detecting levels of the polypeptide of the present invention in cells and tissues, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for detecting over-expression of the polypeptide compared to normal control tissue samples may be
10 used to detect the presence of an infection, for example, and to identify the infecting organism. Assay techniques that can be used to determine levels of a polypeptide, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these, ELISAs frequently are preferred. An
15 ELISA assay initially comprises preparing an antibody specific to the polypeptide, preferably a monoclonal antibody. In addition, a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached to a detectable reagent such as radioactive, fluorescent or enzymatic reagent, such as horseradish peroxidase enzyme.

Antibodies

20 The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. The present invention includes, for example, monoclonal and polyclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the
25 product of an Fab expression library.

 Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a
30 sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

 For preparation of monoclonal antibodies, any technique known in the art which provides antibodies produced by continuous cell line cultures can be used.
35 Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et*

al., pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985); U.S. Patent No. 5,545,403; U.S. Patent No. 5,545,405; U.S. Patent No. 5,654,403; U.S. Patent No. 5,792,838; U.S. Patent No. 5,316,938; U.S. Patent No. 5,633,162; U.S. Patent No. 5,644,036; U.S. Patent No. 5,858,725.

5 Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

10 Alternatively, phage display technology could be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Fbp or from naive libraries (McCafferty, J. et al., (1990), *Nature* **348**, 552-554; Marks, J. et al., (1992) *Biotechnology* **10**, 779-783). The affinity of these
15 antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) *Nature* **352**, 624-628).

 If two antigen binding domains are present, each domain may be directed against a different epitope - termed 'bispecific' antibodies.

 The above-described antibodies may be employed to isolate or to identify
20 clones expressing the polypeptide or purify the polypeptide of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography.

 Thus, among others, antibodies against the polypeptide of the present invention may be employed to inhibit and/or treat infections, particularly bacterial
25 infections and especially infections arising from *S. epidermidis*.

 Polypeptide derivatives include antigenically, epitopically or immunologically equivalent derivatives which form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized
30 by certain antibodies which, when raised to the protein or polypeptide according to the present invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to
35 interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof can be used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by
5 conjugation, with an immunogenic carrier protein, for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively, a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof, may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a
10 carrier.

Preferably the antibody or derivative thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized," wherein the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted
15 into a human monoclonal antibody, for example as described in Jones, P. et al. (1986), *Nature* 321, 522-525 or Tempest et al., (1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscle (Wolff et al., *Hum Mol Genet* 1992, 1:363, Manthorpe et al.,
20 *Hum. Gene Ther.* 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem* 1989:264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA*
25 *Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS* 1984:81,5849).

Binding molecules and assays

This invention also provides a method for identification of molecules, such as binding molecules, that bind to the polypeptide of the present invention. Genes
30 encoding proteins that bind to the polypeptide can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., *Current Protocols in Immunology* 1(2): Chapter 5 (1991). Also, a labeled ligand can be photoaffinity linked to a cell-extract. Polypeptides of
35 the invention also can be used to assess the binding capacity of a binding molecule, in cells or in cell-free preparations.

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics.

5 The invention further provides a complex of a polypeptide and a binding molecule which comprises a polypeptide as described herein and a binding molecule capable of modulating the activity of the polypeptide. A complex of this kind can arise *in vivo* upon administration to a patient of a binding molecule as described herein.

10 Antagonists and agonists - assays and molecules

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the function of polypeptides or polynucleotides of the present invention, such as its interaction with a binding molecule. The method of screening may involve high-throughput.

15 For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, may be prepared from a cell that expresses a molecule that binds to the polypeptide of the present invention. The preparation is incubated with labeled polypeptide in the absence or the presence of a candidate molecule
20 which may be an agonist or antagonist. The ability of the candidate molecule to bind the binding molecule is reflected in decreased binding of the labeled ligand. Molecules which bind gratuitously, *i.e.*, without inducing the functional effects of the polypeptide, are most likely to be good antagonists. Molecules that bind well and elicit functional effects that are the same as or closely related to the polypeptide are
25 good agonists.

The functional effects of potential agonists and antagonists may be measured, for instance, by determining activity of a reporter system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect with that of the polypeptide of the present invention or
30 molecules that elicit the same effects as the polypeptide. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in the functional activity of the polypeptide, and binding assays known in the art.

Another example of an assay for antagonists is a competitive assay that
35 combines the polypeptide of the present invention and a potential antagonist with membrane-bound binding molecules, recombinant binding molecules, natural

substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. The polypeptide can be labeled such as by radioactivity or a colorimetric compound, such that the number of polypeptide molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds to the same sites on a binding molecule without inducing functional activity of the polypeptide of the invention.

Potential antagonists include a small molecule which binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

Other potential antagonists include antisense molecules (see Okano, J. *Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules).

Preferred potential antagonists include derivatives of the polypeptide of the invention.

In a particular aspect, the invention provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: i) in the prevention of adhesion of *S. epidermidis* to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; ii) to block protein mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992)); iii) to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins which mediate tissue damage; iv) to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

Each of the DNA coding sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein

upon expression can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The antagonists and agonists may be employed, for instance, to inhibit diseases arising from infection with *Staphylococcus*, especially *S. epidermidis*, such as sepsis and endocarditis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal, which comprises inoculating the individual with the polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody to protect said individual from infection, particularly bacterial infection and most particularly *Staphylococcus* infections. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises, through gene therapy or otherwise, delivering a nucleic acid functionally encoding the polypeptide, or a fragment or a variant thereof, for expressing the polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibodies or a cell mediated T cell response, either cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise.

A further aspect of the invention relates to an immunological composition which, when introduced into a host capable of having induced within it an immunological response, induces an immunological response in such host, wherein the composition comprises recombinant DNA which codes for and expresses an antigen of the polypeptide of the present invention. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

The polypeptide of the invention or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. This fused recombinant protein preferably further comprises

an antigenic co-protein, such as Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilise the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *S. epidermidis*. Such fragments will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. This approach can allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of *S. epidermidis* infection in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused e.g. by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The present invention also includes a vaccine formulation which comprises the immunogenic recombinant protein together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include

suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials, and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include
5 adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in-water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain
10 polypeptides, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions

15 The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or the agonists or antagonists. Thus, the polypeptides of the present invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions
20 comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration.

Kits

25 The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. The ingredient(s) can be present in a useful amount, dosage, formulation or combination. Associated with such
30 container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

Administration

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

5 The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

10 The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. In general, the compositions are administered in an amount of active agent of at least about 10 $\mu\text{g/kg}$ body weight. In most cases they will be administered in one or more doses in an amount not in excess of about 8 mg/kg body weight per day. Preferably, in most cases, dose is from about 10 $\mu\text{g/kg}$ to about 1 mg/kg body weight, daily. For
15 administration particularly to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg and typically around 1 mg/kg. For example, a dose may be 1 mg/kg daily. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity,
20 route of administration, complicating conditions and the like. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such
25 are within the scope of this invention.

 In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

 Alternatively the composition may be formulated for topical application, for
30 example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for
35 example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such

carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

The pharmaceutical composition may be administered in conjunction with an in-dwelling device. In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters, etc.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent *Staphylococcus* wound infections.

Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5µg/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks.

With the indicated dose range, no adverse toxicological effects should be observed with the compounds of the invention which would preclude their administration to suitable individuals.

The antibodies described above may also be used as diagnostic reagents to detect the presence of bacteria containing the protein.

In order to facilitate understanding of the following example certain frequently occurring methods and/or terms are explained in the foregoing glossary.

5 The present invention is further described by the following examples. While illustrating certain specific aspects of the invention, the examples do not portray the limitations or circumscribe the scope of the disclosed invention.

All examples were carried out using routine molecular biology techniques as generally described in standard laboratory manuals, such as Sambrook et al.,
10 *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

EXAMPLES

15 A small insert plasmid library was generated in the minimal sequencing vector pOT2a (O. Hubbard, C. Martin, and M. Palazzolo, unpublished). pOT2a vector was prepared by BstXI digestion of the parent plasmid pOT2a-sacB followed by preparative agarose gel electrophoresis to separate the 1.6 kb vector fragment from a *B. subtilis* sacB gene fragment. To prepare inserts for library
20 construction *S. epidermidis* SR1 strain genomic DNA was sonicated and the resulting random fragments were end-repaired with klenow and T4 polymerase and phosphorylated with T4 polynucleotide kinase. Oligos (5'-CTCTAAAG-3', 5'-CTTTAGAGCACA-3') (SEQ ID NO.:4465) to create BstXI adaptors were annealed and ligated to the blunt-ended fragments. The configuration of the BstXI
25 sites in pOT2a and the sequence of the adaptors allowed a ligation strategy that minimized the recovery of clones without insert (Seed, 1987). DNA samples were electrophoresed on a low-melting-temperature agarose gel and fragments of 3000-4000 bp were isolated and purified. The linearized vector and random DNA fragments were ligated overnight using T4 DNA ligase at 16C and transformed
30 into DH10B competent cells (Life Technologies Inc., Gaithersburg, MD) by electroporation. Transformed bacteria were selected on LB agar plates containing 5% sucrose and 12.5 ug/ml chloramphenicol. Sequencing templates were isolated from single colonies and purified using R.E.A.L. Prep 96 Plasmid Kit (QIAGEN, Chatsworth, CA). Seq01 primer (5'-CACTATAGAACTCGAGCAGCTG-
35 3') (SEQ ID NO.:4466) and seq02 primer (5'-CGACTCACTATAGGGAGACCG-3')

(SEQ ID NO.:4467) were used to generate end-sequence using ABI Prism BigDye Terminators (PE Applied Biosystems, Foster City, CA).

Constructs from the pOT2a library were transformed into POX38 bacteria and selected on LB agar plates containing 12.5 ug/ml chloramphenicol. A single colony from each construct was used to inoculate an overnight culture. These POX38 cultures were mated with a culture of the F-bearing, kanamycin resistant JGM strain by combining the two strains and shaking for 3 hours at 37C without antibiotics. Each successful mating event resulted in the random insertion of a single gamma-delta transposon into the pOT2a construct. This collection of transpositions was captured in the JGM cells by selection of the mated cultures on LB agar plates containing 12.5 ug/ml chloramphenicol and 25 ug/ml kanamycin. A transposon library was created for each of the original pOT2a library constructs by picking 96 individual colonies. A set of two PCR reactions was performed on each of the 96 library members to determine the position of the transposon integration. PM001 primer (5'- CGTTAGAACGCGGCTACA-3') (SEQ ID NO.:4468) and NGDIR primer (5'-GTTCCATTGGCCCTCAAAC-3') (SEQ ID NO.:4469) were used to determine the integration site distance from the left side of the vector and PM002 primer (5'-GCCGATTCATTAATGCAGGT-3') (SEQ ID NO.:4470) and NGDIR primer were used to confirm the integration position by measuring the distance from the right side of the vector. PCR products were electrophoresed in 1X TBE on 1.4% agarose gels. After gel analysis, a subset of transposon clones was selected for sequencing based upon obtaining an integration site about every 300 bp along the full length of the pOT2a insert. Sequencing templates were purified using R.E.A.L. Prep 96 Plasmid Kit (QIAGEN, Chatsworth, CA). M21 primer (5'-GTAAAACGACGGCCAGT-3') (SEQ ID NO.:4471) and rev primer (5'-CAGGAAACAGCTATGAC-3') (SEQ ID NO.:4472) were used to generate internal sequence using ABI Prism BigDye Terminators (PE Applied Biosystems, Foster City, CA).

The sequences, including ORFs (nucleic acid sequences within SEQ ID NOs 1-3334) and non-ORFs (SEQ ID NOs 3335-4464) are set forth in the Sequence Listing. The non-ORF regions may be particularly useful as diagnostic sequences. The ribosomal RNA genes may also be useful to distinguish between species. Also, intergenic regions generally may be useful diagnostics to establish genus and species of an unidentified microbe, as there may be less selective pressure to maintain fidelity of the sequences in these intergenic regions as compared to intragenic regions.

About 26 different isolates of *S. epidermidis* have been submitted to ATCC listed in their on-line catalog, listed below:

- 1: ATCC Number: 146 Organism: *Staphylococcus epidermidis*
- 2: ATCC Number: 33501 Organism: *Staphylococcus epidermidis*
- 5 3: ATCC Number: 49741 Organism: *Staphylococcus epidermidis*
- 4: ATCC Number: 51625 Organism: *Staphylococcus epidermidis*
- 5: ATCC Number: 29997 Organism: *Staphylococcus epidermidis*
- 6: ATCC Number: 19654 Organism: *Staphylococcus epidermidis*
- 7: ATCC Number: 14389 Organism: *Staphylococcus* sp. deposit
- 10 8: ATCC Number: 14852 Organism: *Staphylococcus epidermidis*
- 9: ATCC Number: 49134 Organism: *Staphylococcus epidermidis*
- 10: ATCC Number: 13518 Organism: *Staphylococcus epidermidis*
- 11: ATCC Number: 9491 Organism: *Staphylococcus epidermidis*
- 12: ATCC Number: 35547 Organism: *Staphylococcus epidermidis*
- 15 13: ATCC Number: 35984 Organism: *Staphylococcus epidermidis*
- 14: ATCC Number: 35983 Organism: *Staphylococcus epidermidis*
- 15: ATCC Number: 700296 Organism: *Staphylococcus epidermidis*
- 16: ATCC Number: 49461 Organism: *Staphylococcus epidermidis*
- 17: ATCC Number: 29641 Organism: *Staphylococcus epidermidis*
- 20 18: ATCC Number: 29887 Organism: *Staphylococcus epidermidis*
- 19: ATCC Number: 29886 Organism: *Staphylococcus epidermidis*
- 20: ATCC Number: 55133 Organism: *Staphylococcus epidermidis*
- 21: ATCC Number: 27626 Organism: *Staphylococcus* sp. deposit
- 22: ATCC Number: 31874 Organism: *Staphylococcus epidermidis*
- 25 23: ATCC Number: 14990 Organism: *Staphylococcus epidermidis*
- 24: ATCC Number: 155 Organism: *Staphylococcus* sp. deposit
- 25: ATCC Number: 155-U Organism: *Staphylococcus* sp. deposit
- 26: ATCC Number: 12228 Organism: *Staphylococcus epidermidis*

30 Throughout this application, various publications are referenced. These publications are hereby incorporated by reference in their entirety.

While the invention has been described with respect to certain specific embodiments, it will be appreciated that many modifications and changes may be made by those skilled in the art without departing from the spirit of the invention.

35 It is intended, therefore, by the appended claims, to cover all such modification and changes as fall within the true spirit and scope of the invention.

What is claimed is:

1. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide having at least a 70% identity to a polypeptide set forth in the Sequence Listing;

(b) a polynucleotide which is complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a) or (b).

2. The polynucleotide of Claim 1 wherein the polypeptide has at least 80% identity to the polypeptide set forth in the Sequence Listing.

3. The polypeptide of Claim 2 wherein the polypeptide has at least 90% identity to the polypeptide set forth in the Sequence Listing.

4. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

5. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

6. The polynucleotide of Claim 4 wherein the polynucleotide has at least 80% identity to a polynucleotide set forth in the Sequence Listing.

7. The polynucleotide of Claim 6 wherein the polynucleotide has at least 90% identity to a polynucleotide set forth in the Sequence Listing.

8. The polynucleotide of Claim 4 comprising a polynucleotide set forth in the Sequence Listing.

9. The polynucleotide of Claim 4 comprising a polynucleotide set forth in the Sequence Listing as any of an odd-numbered SEQ ID Nos 1-3334.

10. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide contained in any of ATCC Deposit Nos. 146; 33501; 49741; 51625; 29997; 19654; 14389; 14852; 9134; 13518; 9491; 35547; 35984; 35983; 700296; 49461; 29641; 29887; 29886; 55133; 27626; 31874; 14990; 155; 155-U; 12228 and substantially encoding the polypeptide comprising amino acids 1 to 416 of SEQ ID NO:2;

(b) a polynucleotide complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a) or (b).

11. A vector comprising the DNA of Claim 4.
12. A host cell comprising the vector of Claim 11.
13. A process for producing a *S. epidermidis* polypeptide comprising expressing from the host cell of Claim 12 a polypeptide encoded by said DNA.
14. A process for producing a cell which expresses a *S. epidermidis* polypeptide comprising transforming or transfecting the cell with the vector of Claim 11 such that the cell expresses the polypeptide encoded by the DNA contained in the vector.
15. A polypeptide comprising an amino acid sequence which is at least 70% identical to a polypeptide set forth in the Sequence Listing.
16. A polypeptide comprising an amino acid sequence which is at least 80% identical to a polypeptide set forth in the Sequence Listing.
17. A polypeptide comprising an amino acid sequence which is at least 90% identical to a polypeptide set forth in the Sequence Listing.
18. A polypeptide comprising an amino acid sequence as set forth in the Sequence Listing.
19. An antibody against the polypeptide of claim 18.
20. An antagonist which reduces or inhibits the activity of the polypeptide of claim 18.
21. A method for the treatment of an individual having need to reduce or inhibit the activity of the polypeptide of Claim 18 comprising administering to the individual a therapeutically effective amount of the antagonist of Claim 20.
22. A complex of a polypeptide and a binding molecule which comprises the polypeptide of Claim 18 and a binding molecule that is capable of antagonising the activity of the polypeptide.
23. A process for diagnosing in a subject a disease related to expression of the polypeptide of claim 18 comprising detecting the presence in the subject of a nucleic acid sequence encoding said polypeptide.
24. A diagnostic process comprising detecting the presence of the polypeptide of claim 18 in a sample derived from a subject.
25. A method for identifying compounds capable of inhibiting the activity of the polypeptide of claim 18 comprising:
 - (a) contacting a cell expressing the polypeptide on the surface thereof with a selected compound, under conditions to permit binding to the polypeptide in the presence of a component capable of providing a detectable signal in response to the binding of the compound to said polypeptide; and

(b) detecting the presence or absence of a signal generated in response to the binding of the compound to the polypeptide,
the presence of a signal indicating a compound capable of inhibiting the activity of the polypeptide.

26. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of Claim 15, or a fragment or variant thereof, adequate to protect said animal against infection from *S. epidermidis*.

27. A method of inducing an immunological response in a mammal which comprises delivering a gene encoding the polypeptide of Claim 15, or a fragment or variant thereof, and obtaining expression of the gene *in vivo* in order to induce an immunological response to produce antibody to protect said animal against infection from *S. epidermidis*.

28. An immunological composition comprising a DNA capable of expressing a polypeptide of Claim 15 which, when introduced into a mammal, induces an immunological response in the mammal, and a pharmaceutically acceptable carrier.

29. An immunological composition comprising a polypeptide of claim 15 which, when introduced into a mammal, induces an immunological response in the mammal, and a pharmaceutically acceptable carrier.

SEQUENCE LISTING

<110> Kimmerle, Bill

<120> STAPHYLOCOCCUS EPIDERMIS NUCLEIC ACIDS

Sequence 1

Contig_0440_pos_11083_0,

putative peptide of unknown function

atgcaagctttacgagataaagtaggccaacaaaataacgttcaccaacaaagtaattat
ttcaatgaagatgaacaacaaaacataactatgataattctgtacaagccggtcaaact
attattgataaaacttcagatccaatcatgaacaaaatgaaattgagcaggctattaat
caaatcaatacagactcaaacagcggttaagtggagaaaaataattacacactgaccaagaa
agcacaatatagacaaatagaaggtttatctagtttgaacacagctcaaatcaacgccgaa
aaagatttagtcaatcaagctaaaaacaagaacagatgttgctcaaaagttagctacagct
aaagaaataaattctgctatgagtaatttaagagatggcattcaaaataaagaggacatc
aaacgttagcagtgcatatatcaacgcagatccgactaaagttacagcttacgatcaagca
ctacagaacgcagaaaaatatcatcaatgccacaccaaacgtagagcttaataaagctaca
attgaacaagcgctatcacgcggttaacaagcacaacaagatcttgatggtggtcaacaa
ttagctaatgctaaacaacaagctacacaaactgtcaatgggttaaatagcttaaatgac
ggtcaaaagcggtgaattaaatctattaatttaattcagctaatcccgtacaaaagtlaca
gaagaattaaacaaagcaactgaatcgaacctatgcgatggaagctttaagaaacagtggt
caaaacggtgatcaagtaaaacaaagtaacaattatgtcaatgaagatcaacctgaacag
cacaattatgataatgctgtcaatgaagctcaagctacaatcaacaacaatgctcaacct
gttctagacaaatttagctatagaacggtttaactcaaaactgttaacactacaaaagatgca
ttacatggttactcaaaaactgatacaagaccaacaagctgctgaaactggaatacgtggt
ttaacgagctctcaatgaacctcagaaaaatgctgaagtagctaaagtaactgcagcaaca
acacgtgatgaagtgaagaaatattcgtcaagaagcaacaacattagatactgcaatgctt
ggtttacgtaaaagcattaaagataaaaacgataactaaaaatagtagtaaatatattaat
gaggatcatgaccaacaacaagcttatgacaatgctgtaataatg

Sequence 2

MQALRDKVGQQNNVHQSSNYFNEDEQPKHNYDNSVQAGQTIIDKLQDPIMNKNEIEQAIN
QINTTQTALSGENKLHTDQESTNRQIEGLSSLNTAQINAEKDLVNQAKTRTDVAQKLATA
KEINSAMSNLRDGIQNKEDIKRSSAYINADPTKVTAYDQALQNAENIINATPNVELNKAT
IEQALSRVQQQQDLDGVQQLANAKQQATQTVNGLNSLNDGQKRELNLLINSANTRTKVQ
EELNKATESNHAMEALRNSVQNVQVQKQSNNYVNEDQPEQHNYDNAVNEAQATINNNAQP
VLDKLAIERLTQTVNTTKDALHGTQKLIQDQQAETGIRGLTSLNEPQKNAEVAKVTAAT
TRDEVNRIRQEATTLDTAMLGLRKSIDKNDTKNSSKYINEDHDQQQAYDNAVNNX

Sequence 3

Contig_0440_pos_9509_8859,

putative peptide of unknown function

atgaatcgtattgcccatagttatgggtttacatgatacatagttttgtgacatcaact
gcaattattttctcattaaatgatcgtactagtacgaggttgattcgtattcgcgaacgt
acaaccgatcttgagaaaattgctttaaccaatagcctatctcgtaaaatttcgagtaag
caacttacaattgacgaagcaaaaagtgaattactgcaacttaaacgtgcgtctcttcag
tattcttttcttaacaaatctcattgctgcctttgtagcttggtgtttttcttattcatg
tttgggtggcgtagcttccgacgcttgattgcatgcctagcgggtggcatagctttttta
acgttttagtttcgtgcaaaaatatatacaaatataattcttttcagagtttgtagcatct
gctgttggttattagttatgcagcaatattcactaaactaggtatagctaaaaatcaagac
attattactattgcaagtgtcatgcctctcgttcccggtattttgattactaacgctatt
cgtgacttacttgccggagagttacttgctggtatgtcacgtggtgttgaagctgcttta
actgcatttgctattgggtgcaggagtagctattgtattactattattataa

Sequence 4

MNRIAHSYGLHDTYSFVTSTAIIFSLNDRSTRLIRIRERTTDLEKIALTNSLSRKISSK
 QLTIDEAKSELLQLKRASLQYSFLTNLIAAFVACGFFLFMFVGGVASDAWIACLAGGIAFL
 TFSFVQKYIQIKFFSEFVASAVVISIAAIFTKLGIKNQDIITIASVMPLVPGILITNAI
 RDLLAGELLAGMSRGVEAALTAFAIGAGVAIVLLLL* .

5

Sequence 5

Contig_0440_pos_8843_8349,

putative peptide of unknown function

atgtttatctgctttcactttatcattagtttcattgccacagtccttttttcaatt
 10 atatttaattgcacctaataattgctattagcttggtgatttggaggctgttgcttg
 acaatatatcagatgacagtaggtatggatttaggttaaagttggcgcttcatttttagga
 agtctaattattagattaatgagtcatacaatgagtagacgggtacaagcaacctgttatt
 atatttatcgtccccggcattatacctctcgttccaggtggcgagcatatgaagctaca
 agatttttagtatcaataattatacgaatgcagttaatacttttttagaggttaacatta
 15 atttctggtgcaattgcattcgggtatacttctgatatgaaatagctctattacattttattca
 cgcatcaagcaatcttatggtaaaatcaagggtaaaacttataaaaaatcctataatatg
 aataatagagtataa

Sequence 6

20 MFIYLFHFIISFIATVLFSSIIFNAPKLLLLACGFGVAVAWTIYQMTVGMDLGKVGASFLG
 SLILGLMSHTMSRRYKQPVIIIFIVPGIIPLVPGGAAYEATRFLVSNNYTNVNTFLEVTL
 ISGAIAFGILVSEIVYIIYSRIKQSYGKIKGKTYKKSYNMNNRV*

Sequence 7

25 Contig_0440_pos_8175_7090,

is similar to (with p-value 0.0e+00)

>sp:sp|P55179|PEPT_BACSU PEPTIDASE T (EC 3.4.11.-) (AMINOTRI
 PEPTIDASE) (TRIPEPTIDASE). >gp:gp|X99339|BSGALE_6 B.subtilis
 orfs 1,2,3,4, pepT and galE genes. NID: gl429253. >gp:gp|Z9
 30 9123|BSUB0020_187 Bacillus subtilis complete genome (section
 20 of 21): from 3798401 to 4010550. NID: g2636240. >gp:gp|D
 83026|D83026_30 Bacillus subtilis genome sequence covering l
 ic-cel region. NID: gl783231.

atggatgaacatgggttacttatttgcactcgaagcaatattaattataatgtacct
 35 actgtcgggttttttagcacatgtagacacttcaccagatttcaatgcttctcatgtaaat
 ccgcaaatcattgaagcctataatgggcaacctatcaaacttggtgaatctcagcgatc
 ttagatcctgatgtttttcctgaattaaataaagttgtgggtcatacactaatggtgaca
 gatggtacatctctactaggcgccgatgataaagcaggtgttgtagaaataatggaagg
 ataaagtatttaattgatcatcctgacattaaacacggtacaattcgagttggctttaca
 40 cccgatgaagaattggacgaggcccgcatcaatttgatgtagtcgatttaattgcagat
 tttgcataatacaatggatggcagtcatttaggagaactacaattcgaaagtttcaatgcg
 gcagaggttaactgtcacttgccatggtgtaaacgttcacccaggttcagctaaaaatgcc
 atgggttaattgcaattagtttaggtcaacagtttaataagtttacttcctcacatgaagt
 cctgaaagaactgaaggatacgaagggttctatcatttaattgaattttacaggtaattgt
 45 gaaaaagcaactctacaatatattatcgcgaccatgacaaagaacagtttgagctacgt
 aaaaaacgcattgatggaattcgtgatgatttaattgttcattataatcattttccaatt
 aaagtagatgtgcatgaccaatattttaacatggcagaaaaaattgaacctttgaaacac
 atcattgatatacctaaacgtgtctttgaggcttttagacatcgtaacctaactgaacct
 attcgaggtggtacagatggatcacaattatcttttatggggttacctacacctaatt
 50 ttactggttggtgcaatttccacggtccttttgaaatcgcttctatcgatgtaattgaa
 aaggctgttcattgtgtcgttggtattgtcgaagaagtagcaaacagccatcaatcttat
 aataa

55 Sequence 8

MDEHGYLFATLESNINYNVPTVGFLAHVDTSDFNASHVNPQII EAYNGQPIKLGESQRI
 LDPDVFPENKVVGHITLMVTDGTSLLGADDKAGVVEIMEGIKYLIDHPDIKHGTIRVGFT
 PDEEIGRGPHQFDVSRFNADFAYTMDGSQLGELQFESFNAAEVTVTCHGVNVHPSAKNA
 MVNAISLGQQFNSLLPSHEVPERTEGYEGFYHLMNFTGNVEKATLQYIIRDHDKEQFELR

KKRMEIRDDINVHYNHFPIKVDVHDQYFNMAEKIEPLKHIIDIPKRVFEALDIVPNTEP
IRGGTDGSQLSFMGLPTPNIFTGCGNFHGPFEYASIDVMEKAVHVVVGIAQEVANSHQSY
K*

5 Sequence 9

Contig_0440_pos_4334_3330,
is similar to (with p-value 0.0e+00)
>sp:sp|P37253|ILVC_BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1
.1.86) (ACETOHYDROXY-ACID ISOMEROREDUCTASE) (ALPHA-KETO-BETA
10 -HYDROXYLACIL REDUCTOISOMERASE). >gp:gp|L03181|BACILNB_3 Bac
illus subtilis ilvB, ilvN and ilvC genes, complete ilv-leu o
peron. NID: g143090. >gp:gp|Z99118|BSUB0015_94 Bacillus subt
ilis complete genome (section 15 of 21): from 2795131 to 301
3540. NID: g2635200. >gp:gp|Z75208|BSZ75208_74 B.subtilis ge
15 nomic sequence 89009bp. NID: g1769994.
atgacaaaagtatattacgacgaacagtaactcaggatgcattacaaggtaaaaaaatt
gctgtcattgggttatggctcacaaggacatgcacatgcacaaaatttaaggacaatggt
tatgatgtatgcattggcctgctccaggacgatcatttaataaagctaaagaagatgga
tttgatgtttatcacgtaagtgaagcaacacaacagcagatgtatgtatggtactattg
20 ccttagaaattcaaggtgaagtatatatacaaggaaattaaacatatattagaaaaagga
aatgcttttagcattcgacacgggttttaatatccatttcagtgttatcgaaccacctagt
gatgtcgatgtcttttagtagcacctaaaggacctggctcatttagtttagacgtacattt
gttgaaggaagtgcctgaccagcattatttggtgttcaacaagatgctacaggccaagct
agaaaacattgctttaagctacgcaaaaggcattggtgctactcgtgccggggtcattgaa
25 acgacatttaaaagaagaaactgaaacagatttattcgggtgaacaagctgtactttgtgga
ggagtttccaaatttaattcagagtggattcgaaacacttgggaagcaggttaccacact
gaattagcattattttgaagtccttacacgaaatgaaatttaattggtgatttaattgtatgaa
ggcgggaatggaaaatgtccgttattctatctctaactgctgaatttggcgactatggt
tctggaccaagagtaattacacctaattgtaaagaaaatgaaaaaagtacttgaagat
30 attcaaaatggtaacttttagccgtagatttggtgaagataacaaaaatggctttaagaa
ttctatcaattacgtgaagatcaacatggtcatcaattgaacaagttggacgtgaatta
agagaaatgatgccattcattaaatctaaaagtattgaaaaataa

Sequence 10

35 MTKVYYDETVTQDALQGKKIAVIGYGSQGHAAQNLDNGYDVVIGLRPGRSFNKAKEDG
FDVYTVSEATQQADVVMVLLPDEIQGEVYNKEIKPYLEKGNALAFHGFNIHFSVIEPPS
DVDVFLVAPKPGHLLVRRTFVEGSAPALFGVQQDATGQARNIALSYAKGIGATRAGVIE
TTFKEETETDLFGEQAVLCGGVSKLIQSGFETLVEAGYQPELAYFEVLHEMKLIVDLMEY
GGMENVRYSISNTAEFGDYVSGPRVITPNVKENMKKVLIEDIQNGNFSRRFVEDNKNNGFKE
40 FYQLREDQHGHIQEVGRELREMPFIKSKSIEK*

Sequence 11

45 Contig_0440_pos_3316_1772,
is similar to (with p-value 0.0e+00)
>gp:gp|U92974|LLU92974_14 Lactococcus lactis unknown gene, p
artial cds, and HisC (hisC), unknown, HisG (hisG), unknown,
HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF),
50 HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB),
LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB),
IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR)
genes, complete cds. NID: g2565137.
gtgtttttaatggaagaacatatattcaaatttttgatacaacacttagagatggtgaacaa
55 acgccaggagtcatttttacttttgatgaaagattaaaaattgccaagcaactagaaaaa
tggggagtagatgtactagaagcaggttttctgcttctagtactggttagctttaaatca
gtagaagctatagctaaaactttgactacaacagcagtggtgtggttttagctagatgtaaa
aaatctgatattgatgctgtatgaagccactaaagaagctgtaaaacctcaagtacat
gtattcattgcaacctcccctattcatttagaacataaattaaaaatgactcaagatgaa

gttttaacatcaataaaaagaacacggtttcttatgcaaaacaattttttgaagtcgtacaa
 ttctctccagaggatgcaacaagaactgaaattccatttttaattgaatgtgttcaaact
 gcgattaacgcaggagccacaattatcaacatccctgatacagttggatttagttatcct
 acagagtatggcgaaatttttaacaattaacacaggccggttaagtcaaatttctaaaatt
 5 atcttttagtgcacattgtcatgatgatctcggaatggcagttgctaatagttttagcagct
 attgaaggtggagctagacgtattgaaggtaccgtgaacggtattgggtgaaagagcagga
 aatgcctcacttgaagaagtcgctttggctttatatgtaaggaaggaccactatggctctt
 gaatctcaaattaaccttgaagaaactaagaaaacatctgacttaatttcaagatatgct
 ggtatccgtgtacctagaataaagctatagtcggtcaaaatgcatttagtcatgaatcc
 10 ggaattcaccaagacggtgtccttaaacatcgtgaaacctatgaaatcatgacacctcaa
 cttgtaggtgtgaatacaacagaattgccactaggtaaattgtctggtaaacaatgcattt
 gccgaaaagcttaagctctgggatatgaaattaaattggaagatcaagttacattattt
 aaacaattttaaagaattgcccagataagaaaaaaatgtatccgatagagatatcatgcg
 attatacatggctccgaacatgaacacaatgctatttttcaacttgataacttacaactt
 15 caatacgtatctaaaggtctacaaagtgcagtagtagttataaaggaaagaaacggacaa
 gttaacaagattcaagatttgaacgggttcaattgttgcaatttataatgctgttgac
 cgaattttcaagaaagacgcagaattaattgattatcgtattgattctgtaacagaaggt
 actgatgctcaagcagaagtacatgtacgaatcattattaatcatattgaagtgcaggc
 ataggtatagaccacgatataattaaaagcttcatgtaaagcatatatcgatgctcatgct
 20 aatatatttcagaatatgagttgaaagaaggtatacgtacatga

Sequence 12

VFLMEEHIQIFDFTTLRDGEQTPGVNFTFDERLKI AKQLEKWGV DVLEAGFPASSTGSFKS
 VEAIAKTLTTTAVCGLARCKKSDIDAVYEATKEAVKPQVHVFIATSPIHLEHKLKMTQDE
 25 VLTSIKEHVS YAKQFFE VVQFSPEDATREI PFLIECVQTAINAGATI INIPD TVGFSYP
 TEYGEI FQKLTQAVKSN SKI IFSAHCHDDLGM AVANSLAAIEGGARRIEGTVNGIGERAG
 NASLEEVALALYVRKDHYGLESQINLEETKKTSD LISRYAGIRVPRNKAI VGNQAFSHES
 GIHQDGV LKHRETYEIMTPQLVGVNTTELPLGKLSGKHAF AEKLKALGYEIKLEDQVTLF
 KQFKEIADK KKNVSDRDIHAI IHGSEHEHNAIFQLDNLQLQYVSKGLQSAVVVIKERNQG
 30 VKQDSSIGTGSIVAIYNAVDRIFKKDAELIDYRIDSVTEGTDQAQAEVHVRI INHIEVTG
 IGIDHDILKASCKAYIDAHA KYISEYELKEGIRT*

Sequence 13

35 Contig_0440_pos_1280_732,
 is similar to (with p-value 4.0e-48)
 >sp:sp|Q02143|LEU3_LACLA 3-ISOPROPYLMALATE DEHYDROGENASE (EC
 1.1.1.85) (BETA-IPM DEHYDROGENASE) (IMDH) (3-IPM-DH). >pir:
 pir|S35133|S35133 probable 3-isopropylmalate dehydrogenase (EC
 40 1.1.1.85) - Lactococcus lactis subsp. lactis >gp:gp|U9297
 4|LLU92974_15 Lactococcus lactis unknown gene, partial cds,
 and HisC (hisC), unknown, HisG (hisG), unknown, HisB (hisB),
 unknown, HisH (hish), HisA (hisA), HisF (hisF), HisIE (hisI
 E), unknown, unknown, LeuA (leuA), LeuB (leuB), LeuC (leuC),
 45 LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB), IlvN, IlvC
 (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR) genes, comp
 lete cds. NID: g2565137.
 gtgagaatcgctttttaatcttgcaaatcgtagacgtaaaaaattaacttctgttgataag
 gaaaacggttctatctctagtaagttatggcgacaaatagtaaacgatgtaaaaaaggat
 50 tatccagaagtagaagttaatcatatgctagttgatgctttagcatgcatctgatcact
 caacctacgcaatttgatgtgattgtaacagagaatcttttcggagatattttaagtgat
 gaagcatctgttataccagggtctctaggtcttttccatcagctagttttgggtcaaaca
 ggtacacgtctttatgaaccaattcatggttcagcaccagatatagctaataagataaa
 gcgaatccatttggtatggttctatctttagcactttgcttaagagaaagttaaatcaa
 55 aatgatgctgctaacgaacttgagtcgaattgtttactgctttattcaatctaataagaca
 actgcagatttaggtggacaatatcgaacttcagaaatttttaattgcttaagaaaaa
 tatctataa

Sequence 14

VRIAFNLANRRRKKLTSVDKENVLSSSKLWRQIVNDVKKDYPEVEVNHMLVDACSMHLIT
 QPTQFDVIVTENLFGDILSDEASVIPGSLGLSPSASFGQTGTRLYEPIHGSAPDIANEDK
 ANPFGMVLSLALCLRESLNQNDAAANELESIVYSFIQSNKTTADLGGQYRTSEIFKLLKEK
 YL*

5

Sequence 15

Contig_0440_pos_716_354,

is similar to (with p-value 1.0e-32)

>pir:pir|D36889|D36889 3-isopropylmalate dehydratase (EC 4.2
 10 .1.33) chain leuC - Lactococcus lactis subsp. lactis (strain
 IL1403)

atgggtcaaacactgtttgataaagtatggaaaaaacatgtgcttcattggaagaaggt
 gaaccacaattattatacattgattttacatctcattcatgaagtcacttctcctcaagcg
 tttgaaggacttagaatacaaaatcgtaaacacagaagacctgatctaacccttgcaact
 15 ttagatcataacgttcccaattgatatttttaataaaaagatgaaattgctaataaa
 caaattacaactttacaacaaaatgctaaggactttggtgtacataattttgatatgtta
 ctcataattgtcttaagtggacttaacgtatatcttatcattcaaacattccaagaatta
 tga

20 Sequence 16

MGQTLFDKVVKKHVLHGKEGEPQLLYIDLHLIHEVTSPQAFEGRLRIQNRKLRRPDLTFAT
 LDHNVPTIDIFNIKDEIANKQITTLQQNAKDFGVHIFDMLLIIVLSGLNVYLIQTFQEL
 *

25

Sequence 17

Contig_0441_pos_5436_6512,

is similar to (with p-value 0.0e+00)

>sp:sp|P39576|ILVE_BACSU PUTATIVE BRANCHED-CHAIN AMINO ACID
 30 AMINOTRANSFERASE (EC 2.6.1.42) (BCAT). >pir:pir|S57763|S5776
 3 amino acid aminotransferase homolog - Bacillus subtilis >g
 p:gp|Z49992|BSCLABCD_6 B.subtilis celA, celB, celC, celD and
 ywaA genes. NID: g895746. >gp:gp|Z99123|BSUB0020_150 Bacil
 35 lus subtilis complete genome (section 20 of 21): from 379840
 1 to 4010550. NID: g2636240.

atgtcagaaaaagtaaaattcgaaaaagagagtcctttaaaagaaaaacctgatacagca
 aacttaggatttggacaatatctcacagactatatgttaagtgttgattatgacgctgat
 caaggattggcatgatatgaagattgtgccgtacgcaccatttgaaatttcaccagcagcg
 40 caagggttacattatgggtcaggcagttttgaaggccttaaagcctataaacataatgga
 gaagttgtattatccgcccagatcaaaacttcaaacgtattaataattctttagcacgt
 ttagaaatgccagaagttgatgaagaagcattattagaaggttgaagcagcttatcgac
 gttgaacgagattgggtacctgaaggcgaaggtcaatcgttatatattcgtccttttgta
 tttgctactgaaggtgttttgggtgtacgttcttcacatcaatataaattactaattatt
 45 ttatctccgtcaggcgcttattatggtggtgacacattaaagtcaactaaaatttatgtc
 gaagatgaatatgtacgtgcagtcagtgagggtgtaggtttcgctaaagttgcaggtaac
 tatgctgccagcttactcgcacaaaacacgctaataaattagggttatgaccaagtattg
 tggcttgatggtgtgaacaaaaatatgttgaagaagttggttagtatgaatatctcttc
 gtagaaaatggaaaagtagttacgccagcattaaacggtagtatcttgccctggtatcact
 50 agaaagtcaattattcaattagctgaagatttaggttatgaagttgaagagagaagagtt
 tctatagaagagctgtttaacgcataatgataaaggtgaacttacagaagtatgtggttca
 ggtacagcagctgttatctctcctgtaggtacacttcgctatgaagatagagaaattggt
 attaataacaatgaacctggttaaaatcactcaaaaattatatgacacataactggtatt
 caaagtgggcaaatagaagataaaatcggatggagagtagaagttcctaagtatttaa

55

Sequence 18

MSEKVKFEKRESLKEKPDANLGFQYFTDYMLSDYDADQGWHDMKIVPYAPFEISPA
 QGLHYGQAVFEGKLKAYKHNGEVVLFQPDQNFKRINNSLARLEMPVDEEALLEGKQLID
 VERDWVPEGEGQSLYIRPFVFATEGVLGVRSSHQYKLLIILSPSGAYYGGDTLKSTKIYV

EDEYVRAVRGGVGFAKVAGNYAASLLAQTNANKLGYDQVLWLDGVEQKYVEEVGSMNIFF
 VENGVVTPALNGSILPGITRKSIIQLAEDLGYEVEERRVSIEELFNAYDKGELTEVFGS
 GTAAVISPVGTLRYEDREIVINNNEPGKITQKLYDITYGTGQSGKLEDKYGWRVEVPKY*

5 Sequence 19

Contig_0441_pos_6782_7498,

is similar to (with p-value 1.0e-33)

>pir:pir|S60902|S60902 hypothetical protein 1 - Haemophilus
 influenzae >gp:gp|X78559|HISBCAL_1 H.influenzae DNA for sero
 type b capsulation locus. NID: g471233.

10 atgatttatgctgggtatattagcaggtggtattggttctagaatgggaaatgttccatta
 cccaaacaattttttatcattacaaggaaaacattattattcatacagtagaaaaattt
 ttaatgtataaggactttgatgaaatcatcattgccacgcctcaaaagtggatcaattat
 atgctcgatttgctaaacaattatcaattagacgataagaaaataaaagtaatacaaggc
 15 ggagacgaccgaaatcactctataatgaattatagaaagcattgagcaacataaaaaa
 ttaaatgatgaagatataatcggtacccatgatgcagttaggccatttctaacaatcga
 attattagagagaatgtggaatatgccagtcattatggtgcagtagatacggttgtaaat
 gctggtgatactactatttcttcaaatgatgcacaatttatttctgggattccaataaga
 agtgagatgatatcaaggacagcgcctcaaaacttttaaaataaaagagttaaaggatagc
 20 tatttatcggttaactcaatctcaaaaggaaatattaactgacgcgtgtaaaatactcgta
 gaattgggtaagccagtaaaattagtc aaaggagagttatttaacataaaaaataacaaca
 ccatatgatttaaaagttgcgaattcaattattactggagctggtgataatgattaa

25

Sequence 20

MIYAGILAGGIGSRMGNVPLPKQFLSLQGKPIIIHTVEKFLMYKDFDEIIITPQKWINY
 MLDLLNNYQLDDKKIKVIQGGDDRNHSIMNIIIESIEQHKKLNDEDIIVTHDAVRPFLTNR
 IIRENVEYASQYGAVDTVVNAVDTIISSNDAQFISGIPIRSEMYQGQTPQTFKIKELKDS
 30 YLSLTQSQKEILTDACKILVELGKPKVLKVGELFNIIKITPPYDLKVANSIITGAVDND*

Sequence 21

Contig_0441_pos_7863_8522,

putative peptide of unknown function

35 atgaaacctgatagagtagtcactttaccacaagaaattgatttgagtgtagcgtcgat
 actgaattagttacagtttagtggtcatgcaatagatcgttttcaatcaaaagcaatacca
 caatttgaatcattggggatattggggagacggtaatttagggatatcactgcagtttta
 ctaaaaaaactatatcctactacaaaaataatagtttttgaaaaacattatataaatta
 agccggtttttcattttagatgaaattattcaaatgacaatttcctcaacatatcaaa
 40 attgatcatgcatttgaatgcgtgggtggttaagggaagtcagcaggctattgaacaaatt
 attaatattattaatcctgaagggagtagcgtttattaggagtgagcgaactgcctata
 caagtgaataacaagaatggtttttagaaaaaggtttaactataattggtagtagcagaagt
 ggtttaaaggattttgaaaaaactattgaattgtatcgtaaatatcctgaagttcttaaat
 caattagcattacttaaaaggtaagaatttgaaataaataccatagaagatctcattaca
 45 gcgttcgaatatgatatttctaacgcattggggaaaaacagttttaaatggaatatttaa

Sequence 22

MKPDRVVTLPQEIDLVSASYTELVTVSVHAIDRFQSKAIPQFESLGIWGDGNLGYITAVL
 50 LKKLYPTTKIIIVFGKTLKLSRFSFVDEIIQIDNIPQHIKIDHAFECVGGKGSQQAIEQI
 INIINPEGSIALLGVELPIQVNTRMVLEKGLTIIGSSRSGLKDFEKTIELYRKYPEVLN
 QLALLKGKEFEINTIEDLITAFEYDISNAWGKTVLKWNI*

Sequence 23

55 Contig_0441_pos_8539_9987,

is similar to (with p-value 5.0e-32)

>pir:pir|S49240|S49240 hypothetical protein 3 - Haemophilus
 influenzae >gp:gp|Z37516|HIACAPIID_3 Haemophilus influenzae
 serotype a capsulation locus region II DNA. NID: g547510.

atgacaaagcaaaatatatttatagatgacatttattgggaacgtgtccaactcttcgtc
 aaaggacattttgaaggagtaaaacctaacaagaatttccttcttagaaatctaacagaa
 acaaaactattaaatgccaatcatgttaatatcaagggtcaacttttgaggcaagattt
 5 aatattgctatttttagaaaaaggttaatttttaggtacaggcaattatatattaatcaac
 cgacaagaagatgaatatgtctgccaaattaaccccaatttttgatgataaaaaaat
 cagatgacttttagaggagttaagagattacaactcacttgagaccaatcgttacaaaaa
 agttattttatataaaaaagtatggtaaaagtttccaaagatataataacaaagagattaaa
 tcttacgtcattgttccggcaatatcccaagaaattaatgagtttatttttaaagttcaa
 10 tataaatctgaaataaagaaaaataagtaaaacttaagcaattatcatatataattacataaa
 gcttttgaggaaaattagcttcaatgtgagagataaaatatatttgcgtgattttaacatt
 tccaaaacagtatataagaataataaaaaatcatgttttggttacatcagattctagagca
 aatctgcaggaaatttttaatttatatacgaagaatgcttaaacacaattggacaaa
 aaacttgctcattcattctatttttaaacctaataatagcaaataggagatcgttattgat
 aaatataaatttccatatttttttaggaaaactaaatatatcttgggtgatgattatcat
 15 ccgatgatataaaacttcaatttagagaaaaccaagaatagttcaagtatggcatgct
 gtgggtgcttttaagactgtaggatttagtagaactgggaaaaaaggaggacctttcata
 gactctattggacataggaattatagtaaagcttatgttcgtcaaatatgatattctt
 tactatgctgaagcttttgggaattgaagaacatagggttattccaacaggtgttccacgt
 acggatgttttgcgatgaatcttataaaacacgcattaaacaaagttagaaacaaaa
 20 ttaccaattataaaaaataaaaaagtcattcttttgcacctacatttagaggaaatgga
 catcgacagcacactatcctttctttaaaattaattttgcaagattagctagttattgt
 gaagaacatcaagctactgttctgtttaaaatgcatccttttggttagaaataaattaat
 atcccagcaatttatagtaaatatttttagatatttcaaattaccgcgaagtaaatgat
 gtattgttcattacggatattttaatctctgattattcttctttaatctatgaattttcc
 25 agtttttaa

Sequence 24

MTKQNI FIDDIYWERVQLFVKGHFEGVKPTRN FLLRNLTETKLLNANHVNIQGSTFEARF
 NIAILEKGNFLGTGNYILINRQEDEYVCQINPKFLNDKKNQMTLEELRDYNSLETQSLQK
 30 SYLLKKYKGSFQRYNNKEIKSYVIVPAISQEINEFIKVKQYKSEIKKISKLKQLSYILHK
 ALRKISFNVRDKIYLSVFNISKTVYKNNKNHVLFTSDSRANMSGNFKFIYEMLKQQLDK
 KLVIHSIFKPNIANRRSFIDKLKFPYFLGKSKYILVDDYHPMIYKLQFRENQEIQVWHA
 VGAFKTVGFSRTGKKGGPFIDSIHRNYSKAYVSSNNDILYYAEAFGIEHRVIPTGVPR
 TDVLFDES YKTRIKQSLETKLP I IKNKKVILFAPTFRGNGHRTAHYPPFKINFARLAS YC
 35 EEHQATVLFKMHFVRNKLNI PAIYSKYFLDISNYREVNDVLFITDILISDYSSLIYEFS
 SF*

Sequence 25

Contig_0441_pos_1200_13,
 40 is similar to (with p-value 5.0e-90)
 >gp:gp|Y14083|BSY14083_1 Bacillus subtilis chromosomal DNA,
 region 76-78 degrees: between glyB-aprE. NID: g2226224.
 atggaagaagtatacatgtgtttgattgggttcaattagcaagtaataaagaaaagaga
 atgggtgcaattacgacgatatttgcataataccagaactttcttttgaggaaaaacgt
 45 acgcatgattttattgtaaatcaattgagccaattagcatgcaccatagaaacaccagtt
 ggacgtaattggataaaagcaacttttaaaggatctgattcaaatggaccaacgattgca
 ttacgagcagatttcgatgcactacctgttcaagaattaaatgatgtaccctatcgltca
 aaaaataaagggtgcatgcatgcttgtggacatgacggacatacagctattttgcttgga
 gtagctgaaattgttcatgagcatcgctcattttatgaaaggtaattgtgtttttatattc
 50 caatatggtagaggaaattatgccaggtgggtctcaagagatgattgatgatggctgtcta
 cagaatgtcgataaaatatatggcacacacttatggagtggttatccatctgggacaatc
 tattctagacctggagcaataatggcttcaccagatgaatttagtgtgactatatatgga
 aaaggtggtcacggtgcaaaaccacacgaacaatagaccctattgtcattatggctgag
 tttatttttaagtgcacaaaaaataatttctcgaacaattgatccagtaaaaggaagctgtt
 55 ctacttttcggaatgattcaagcaggatcaacagatagtggtatttcagatacagctttt
 tgtaaagggtactgtacgtacttttgacacaaaattacaaagtcattgtcaaaataaaatg
 gataagctcttacaaggtttatctttatcaaacgatattacatatgaattggaatatatt
 aaaggttatttaccagtacacaatcatcaacaatcatatgatgtagtcaaacaagcagct
 aatgattttacatttaagatttaataagagtcagacttaataatgatgattggtaggacttttca

cattaccttaaagtagcacctgggtgcattcttcttaactgggttggtgaataaagacaaa
ggcattactgcacctcatcataatcctcatcttgacattgatgaatcttcattaaaatat
gcagctagtgaatttttaaaaatatttagaaattgaaaatgttttttaa

5 Sequence 26

MEEVIHVDFWFLASNKEKRMVQLRRYLHQYPELSFEEKRTHDFIVNQLSQLACTIETPV
GRNGIKATFKGSDSNGPTIALRADFDALPVQELNDVPYRSKNKGCMHACGHGHTAILLG
VAEIVHEHRHLLKGNVVFIFQYGEIIMPGGSQEMIDDGCLQNVDKIYGTHLWSGYPSTI
YSRPGAIMASPDEFSVTIYGKGGHGAKEPHETIDPIVIMAEFILSAQKIISRTIDPVKEAV
10 LTFGMIQAGSTDSVIPDTAFCKGTVRTFDTKLQSHVQNKMDKLLQGLSLSDITYELEYI
KGYLPVHNHQSYDVVKQAANDLHLRFNESDLMMIGEDFSHYLKVRPGAFFLTGCGNKDK
GITAPHHNPHFDIDESSLKAAASEFLKILEIENVF*

Sequence 27

15 Contig_0442_pos_3158_4903,
is similar to (with p-value 0.0e+00)
>sp:sp|O06446|SECA_STAAU PREPROTEIN TRANSLOCASE SECA SUBUNIT
. >gp:gp|U97062|SAU97062_1 Staphylococcus aureus NCTC 8325 S
ecA (secA) gene, complete cds. NID: g2078389.
20 atgggtggtattgctatacataaagggtgatattgcagaaatgagaacaggtgaagggaaa
acattgactgcaaccatgccgacgtatttgaatgcttagctggttagaggtgtacatggt
attacagtcaatgaatatctatcaagttcacaaagtgaagaaatggctgaactatataac
tatcttggttaactgtaggtttgaactaaatagtaagtcaactgaagaaaaacgtgag
gcttacgcacaagatatcacttatagtagcaataatgaacttgggtttgattatcttaga
gataatatggtgaactatgctgaagagagagtaatgcgtcctctacatttgcaattatt
25 gatgaggtcgattccatattgatcgacgaagcaagaacacctttaattatttctggtgaa
gcggaaaaatctacttctttatatactcaagcaaatgttttgcaaaaatgcttaagcg
gaagatgattataattatgatgaaaaaaccaagctgtacatcttacagaacaaggtgca
gataaagctgaacgtatgttcaaagtagataatctttatgatgttcaaaatgtggaagt
30 attagtcatattaatacagctttaagagctcatgttactttgcaacgcgatgttgattac
atggtcggttgacggtgaagtatttaattgttgaccaatttactggacgtacaatgcctgga
cgtcggttttctgaaaggtttacaccaagcaattgaggctaaagaaggtgtagcaattcaa
aatgagtcataaacgatggcatccattactttccaaaactatttcagaatgtataataag
ttagcggggatgactggtacagcgaaaaccgaagaggaagaatttcgtaatatctataat
35 atgacagttaccctaaattccaacaaacaaacctgttcaacgtaaaagataattcagactta
atztatattagtcataaaaggaaagtttgatgcggtagttgaagatgtttagaaaaacat
aaaaaaggacaaccgtcttactaggtactgttgctgttgagacttctgaatatatttca
aatttactaaaaaaacgtggtgtcagacatgacgtattaaacgctaaaaatcatgaacgc
gaagctgaaatcggttcaaacgcggggcaaaaagggtgcagttacaattgccacaaatag
40 gctggacgtggaacagatattaaacttggtgatggtgttgaagagttaggtggacttgct
gttattggtactgagcgtcatgaatcaagacgtattgatgatcaattacgtggacgttca
ggacgccaaaggtgatagaggagatagtcgtttttacctatctttacaagatgaattaatg
gtacgttttggttcagaacgcttacagaaaatgatgaaccgttttaggaatggatgattca
acgccaatcgagtcgaaaatggtatctcgagctgtagaatcagctcaaaaacgagtagaa
45 ggtaataactttgacgcgcgttaaacgtattctagaatacagatgaagttttacgtaagcaa
cgtgaaattatttataatgagcgtaatgaaatcattgatagtgaaagaaagtctcaagtc
gttaacgcgatgttacgttctacattgcaacgtgacgattaatcattttattaatgaagaa
gacgataatcctgactacacgccatttatcaattacgttaatgatgtgttcttgctgaat
tattaa

50 Sequence 28

MGGIAIHKGDIAEMRTGEGKTLTATMPTYLNALAGRGVHVITVNEYLSSSQSEEMAELYN
YLGLTVGLNLNSKSTEEKREAYAQDITYSTNNELGFDYLRDNMVNYAEERVMRPLHFAII
DEVDSILIDEARTPLIISGEAEKSTSLYTQANVFAKMLKAEDDYNDEKTKAVHLTEQGA
55 DKAERMFKVDNLYDVQNVESHINTALRAHVTLQRDQVDMVVDGEVLIVDQFTGRTPMG
RRFSEGLHQAIEAKEGVAIQNESKTMASITFQNYFRMYNKLAMGTGTAKTEEEFRNIYN
MTVTQIPTNKPVRKDNLDLIYISQKGFDAVVEDVVEKHKKGQPVLLGTVAVETSEYIS
NLLKKRGVRHVDLNAKNHEREAEIVSNAGQKGAVTIATNMAGRGTDIKLG DGVEELGGLA
VIGTERHESRRIDQLRGRSGRQGRGDSRFYLSLQDEL MVRF GSERLQKMMNRLGMDDS

TPIESKMVSRAVESAQKRVEGNFDFARKRILEYDEVLRKQREIIYNERNEIIDSEESSQV
VNAMLRLSTLQRAINHFINEEDNDPDTFPIYVNDVFLNLY*

Sequence 29

5 Contig_0442_pos_6048_7553,
is similar to (with p-value 3.0e-31)
>sp:sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) A
LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TEICHOIC ACID BIOSYN
THESIS PROTEIN E). >pir:pir|S06048|S06048 probable rodD prot
10 ein - Bacillus subtilis >gp:gp|X15200|BSRODC_1 Bacillus subt
ilis rodC operon. NID: g40098. >gp:gp|Z99122|BSUB0019_70 Bac
illus subtilis complete genome (section 19 of 21): from 3597
091 to 3809700. NID: g2636029.
atgatataattctatcggtagaatttaggtaataaattaacaggtatagaacaagctatg
15 atcaatagattaaagctatttaagataatttagtcccaataaactcatattcacatct
tggtcaccacggtttatatatgcatgcacattcggttaaacatcgattcaaaagatattttc
agtctttacgattttctacaagatagattaaactttgagaaaaaacatattgattggata
aattattggcaaaatataatgtaattataccttaaaattcggttgaaaatacgaatgatatt
aaaataacgataaacgacacataaaaatgtatgtgcattttggtgattcaaatatcaa
20 acttttagactatattaaccattttgatatacaacaacgtaaaattcgaagagatttttac
gatacaagaggctttttaagttgtagtagaatttttaacctctcaacaaaagtcgtgatg
gaacaattttttacacctacacaaaaagttaaatttcaaaaatattacaacctgagcac
gaacatcctacggtacaatctatcatttataataacttcacgagacggttcgttttttcaac
gatgaaaatgaacttttagcggttgcaattaatgcgctatatcatttaggagacgtattt
25 ttatgtgataaaaacatcggttacaggcgcaatcattgatcaaaactgatactaaaatacca
gttctcgccggttttccacagtagccacgtaaaaaatataatgatataatcattctgaa
atcaaaacaagcgtataaacctgttttagataacttatcccgatattcaggaatcatagta
tctactgaacaacaaaaaacagatttatctgtaaaaattaataacgttattccattttac
gttataacctgcaggttatattgatacaaatgaatctcatcatagtagtgacaataaacca
30 ttgcctaacaaaatgatattctatcgccggttatttctctgaaaagcaattagatcatcaa
atagaactaatgtctaagctagttccagcatttccaaatttacagttacatttatttggg
tttggttaaagaagaacacacattatcgtaaaattaatagctcaatatcatttagaaaatcac
gtgtttttacgtggttcatttatgattttaaatacaagaaatagagaccgctatttatct
ttattgacaagtaaaatggaagggttcaatttaggtgtacttgaaacgattgctaaaggc
35 gtacctacagtaagttatgataccaaatacggaccttctgagtttaattgttaatacataaa
aatggcgttttttaattgaacaagataataaagaacaactctatcacagcgttaaaaagtta
ttactcgattctaacttaagagaacaattttctaaggaaagattataaacatgcccaata
tttaatgacaaaaatgtttttgatacatggctcactgttttcagaacgttaaaagttaat
ttataa

40 Sequence 30
MIYSIGKNLGNKLTGIEQAMINRLKLFKDNLPNKLIFTSWSPRLYMHASLNIDSKDIF
SLYDFLQDSINFEEKKHIDWYQWQNICNYTLKFVENTNDIKIYDNDTYKMYVHFVDSNYQ
TLDYINHFIDQQRKIRDFYDTRGFLSCSRILTSQQKVMEQFFPTQKVKFQKYNNPEH
45 EHPTVQSIIYNTSRDVRFFNDENELLAFAINALYHLGDFLCKNIVTGPIIDQTDTKIP
VLAVFHSTHVKNINDIYHSEIKQAYKPVLDNLSRYSGIIVSTEQQKTDLVSKINNVIPIY
VIPAGYIDTNESHSSDNKPLPNKMISIGRYSPEKQLDHQIELMSKLVPAFPNLQLHLFG
FGKEETHYRKLIQYHLENHVFLRGFIYDLNQEIETAYLSLLTSKMEGFNLGVLETIAGK
VPTVSYDTKYGPSELIVNHKNGFLIEQDNKEQLYHSVKLLLDNSNLREQFSKESIKHAQI
50 FNDKNVFDTWLTVFRTLKVNLY*

Sequence 31

Contig_0442_pos_10138_10893,
putative peptide of unknown function
55 atgaatctgagtcactgtctgcgtccgaatcactatctgcgtcggagtcgctatctgaat
cggaatcactatctgcgtctgagtcgctatctgaatcggaatcactgtcggagtcgagtc
cactgtctgaatccgagtcactgtctgaatcggaatccgaatcgctgtctgaatccgagtc
cactgtctgaatctgagtcactgtctgcgtccgaatcactatcagaatccgagtcgctgt
ctgcgtccgaatcactgtctgaatctgagtcactgtctgcgtccgaatcactgtcagaat

ccgagtcgctgtctgcatccgaatcactatctgaatccgagtcgctgtctgcatccgaat
 cactatctgaatccgagtcgctatctgcatccgaatcactatctgctgtctgagtcactgt
 ctgctccgaatcactatctgctgtgagtcactgtctgaatccgaatccgaatcactat
 5 cagaatccgagtcgctgtctgagtcagaatcactatctgaatctgagtcactgtctgctg
 cagaatcactatctgctgtgagtcactgtctgaatccgaatccgaatcactatcagaat
 ccgagtcgctgtctgagtcagaatcactatctgaatctgagttactgtctgctgagat
 cgctgtctgctgagtcactgtcgagtcgaatcactatctgctgtgagtcactgt
 ctgaatcatcgctcaaaataaccattatctataactaa

 10 Sequence 32
 MNLSHCLRPNHLYRRSRYLNRNHYLRLSRYLNRNHCRSLSHCLNPSHCLNRRNPNRCLNPS
 HCLNLSHCLRPNHLYQNPSCRCLRPNHCLNLSHCLRPNHQNPSCRCLHPNHLYNPSRCLHPN
 HYLNPSRYLHPNHLYLRLSHCLRPNHLYLRLSHCLNRRNPNHLYQNPSCRCLSQNHLYNLSHCLR
 QNHLYLRLSHCLNRRNPNHLYQNPSCRCLSQNHLYNLSYCLRQNRCLRPSHCRSLNHYLRLSHC
 15 LNHRQNNHYLY*

20 Sequence 33
 Contig_0442_pos_13870_13430,
 is similar to (with p-value 4.0e-18)
 >gp:gp|AJ005645|SAU5645_1 Staphylococcus aureus sdrC gene. N
 ID: g3550591.
 gtgacggataccaatgcgatggttagatagcttcaatcctgatttaaatagttctaagtga
 25 aaagatgtgacaagtaatttacacctaaagtaagtgcagatggtactagagttgatatac
 aattttgctagaagatggcaaatggtaaaaagtataattgtaactcaagcagtgagacca
 acgggaactggaaatgtttataccgaatatgtggttaacaagagatggtactaccaataca
 aatgatttttatcgtggaacgaagtctacaacggtgacttatctcaatggttcttcaaca
 gcacagggggataatcctacatatagcttaggtgactatgtatggttagataaaaaataaa
 30 aacggtgttcaagatgatgatgagaaaggttagcactgagagatccccctcataatttcc
 ccaaagcgtaaacatgtgtga

Sequence 34
 35 VTDTNAMVDSFNPDLNSSNVKDVTSQFTPKVSADGTRVDINFARSMANGKKYIVTQAVRP
 TGTGNVYTEYWLTRDGTNTNDFYRGTKSTVTYLNSSSTAQGDNPYSLGDYVWLDKNK
 NGVQDDDEKGLALRDLIISPKRNHV*

Sequence 35
 40 Contig_0442_pos_10881_9397,
 putative peptide of unknown function
 atggttattttgacgatgattcagacagtgactcagacgcagatagtgattcagactccg
 acagtgactcggacgcagacagcgattctgacgcagacagtaactcagattcagatagtg
 attctgactcagacagcgactcggattctgatagtgattcggattccgattcagacagtg
 actcagacgcagatagtgattctgacgcagacagtgactcagattcagatagtgattctg
 45 actcagacagcgactcggattctgatagtgattcggattccgattcagacagtgactcag
 acgcagatagtgattcggacgcagacagtgactcagacgcagatagtgattcggatgcag
 atagcgactcggattcagatagtgattcggatgcagacagcgactcggattcagatagtg
 attcggatgcagacagcgactcggattctgacagtgattcggacgcagacagtgactcag
 attcagacagtgattcggacgcagacagcgactcggattctgatagtgattcggacgcag
 50 acagtgactcagattcagacagtgactcggattcagacagcgattcggattccgattcag
 acagtgactcggattcagacagtgactcagactccgacagtgattccgattcagatagcg
 actcagacgcagatagtgattccgattcagatagcgactccgacgcagatagtgattcgg
 acgcagacagtgactcagattcagacagtgactcagattcagacagtgattcggacgcag
 actcagactcggactccgacagcgattcagactcagatagtgactcagacgcagacagtg
 55 actcggactcagatagtgattcagatgcagaaagcgattcagactcagatagcgactccg
 attcagacagcgactccgactcagacagtgattcggattcagacagcgattcggactcag
 atagtgactcagacgcagatagtgattccgattcagatagcgactccgattctgatagtg
 actccgattcagatagcgactccgattcagatagtgattcggacgcagacagtgactcgt
 actcagatagtgactccgattcagacagtgattcggattccgatagcgattcggattccg

atagtgactcggattcagacagtgattcggactcagacagcgactccgattcagatagtg
 attccgactcagacagcgattcggattccgatagtgactcggattcagacagtgattcgg
 actcagacagcgattccgattccgatagtgactcggattcagacagtgattcgggctcag
 acagcgattccgattcagacagtgactcggactcagatagtgactccgattcagacagcg
 5 actcggattcttgataaaaatgcaaaagataaattacctgatacaggagcaaatgaagatc
 atgattctaaaggcacattacttggaactttatttgcaggtttag

Sequence 36

MVILTMIQTVTQTQIVIQTPVTVRTQTALITQTVTQIQIVILTQTATRILIVIRIPIQTV
 10 TQTQIVILTQTVTQIQIVILTQTATRILIVIRIPIQTVTQTQIVIRTQTVTQTQIVIRMQ
 IATRIQIVIRMQTATRIQIVIRMQTATRIQIVIRMQTATRIQIVIRMQTATRIQIVIRMQ
 TVTQIQTVTRIQTATRIQIVIRMQTATRIQIVIRMQTATRIQIVIRMQTATRIQIVIRMQ
 TQTVTQIHTVTQIQTVIRTQTVTPTPTAIQTQIVTQTVTQIVIQMQKAIQTQIATP
 IQTATPTQTVIPIQTATRTQIVTQTVIPIQTATPIQIVIPTQTAIRIPIVTRIQTVIR
 15 TQIVTPIQTVIRIPIAIRIPIVTRIQTVIRTQATPIQIVIPTQTAIRIPIVTRIQTVIR
 TQTAIRIPIVTRIQTVIRAQTAIRIQTVTRTQIVTPIQTATRIQIVIPTQTAIRIPIVTRIQTVIR
 MILKAHYLELYLQV*

Sequence 37

20 Contig_0442_pos_9263_7629,
 putative peptide of unknown function
 atgagtatggaaaatcatatagaaagattgattaatcatgttgaaaaaacaatagaaata
 aaagaatatgcttttttaagccttggaatcctaataaaaagccaaagttaaattatta
 aaaaagcctaattaccttagaaggatattactaaagaaattcaaaagtttagacagaaa
 25 acaggagcggtttccttcatgggtaaaaatagacattgttactgaagaagaagttacttta
 ttaaaagatgttaaaagatgaattaaacgcaaaactagaagaatttatattgattttgggtata
 gcttttagatcaatactggaatttatcatttttacctgaagaaataaacactaatgcattt
 attaaaccagtgaaaacagatgggaaaacgaagcttattctatctgaacaaaataataat
 aactatttaagaaagtatacgaaccataaagaaaagtttgcttatgattttatgaaaac
 30 aaagaagtcattaagttcaaaaactaaaggttttatcttagacgaacaaaagatatatgaa
 ttacacgatgaaggtataaaaaaggattaagaaaggtcgattatttacataaagaaata
 gaccaattaattgaaagtggtacatatctcctaggaatatgctatcagatactggaaga
 tatcaatatggtttattttccacatttttgataaagaaatcaatttctataatatattaaga
 catgcttcttcaacttatgcattaatagagggttagattatttaggagaagatttaact
 35 atagtcgaaaaggcaattaactacgttattgagaattatttctatgataatgaaggtgtt
 ggatatatctttgatgatacaaaagatataaacgaaataaaattaggacaaaatgctgcc
 ttatatatttgcggtttgtgaatatttaaaagcataaccccaataagcaatacttatgcgtg
 gcacaaaaagttgctcgaggaattttatcaatgattaatcaagatacatacgaacaact
 catatttttaattatccggatttaactgtgaaagaatcatttagaattatttattatgat
 40 ggtgaagcagcttttagcattattacgcttatatcaccaagatcataatgataaatgggtg
 gaagttgtgaagaagttaatggatcgatttattgaaaaagagttattggcaataccatgat
 cattggcttgggtattgcacgaacgagttagttcaattgtgtccgcaagataaatatttt
 gaatttgggattaagaatgtgaacacctatttggaatatattgaacaacgtgaaacaaca
 ttcccaacatttttagaaatgttaatggcaacctataagcttattcaaaaagcaaaagct
 45 acacatcgtcaagagcttgtgactcagttgattgatgaagaaaaattaatcaatgtgatt
 catacaagagcagaatatcaacgagtaggatttttctatcctgaaattgcaatgtacttc
 aaaaatccaaagcgaattcttggtagtttcttcataaaaacaccatgggtatcgagttcgg
 attgatgatatcgacattatatatcaggctatgttcaatatcaaaaggcacagattaaa
 gatgaaatattatag

Sequence 38

MSMENHIERLINHVEKTIEIKEYAFLSLGKSNIAKVKLLKKPNYLRRDITKEIQKFRQK
 TGAFPSWVKIDIVTEEEVTLFKDVKDELDTQTRRNYIDFGIALDQYWNLSFLPEEINTNAF
 IKPVKTDGKTKLILSEQNINNYLRKYTNHKKKFAYDFYENKEVIKFKTKGFIIDEQKIYE
 55 LHDEGYKKGLRKVDYLHKEIDQLIESGTYFLGNMMLSDTGRYQYGYFPHFDKEINFYNILR
 HASSTYALIEGLDYLGEDLTIVEKAINVYIENYFYDNEGVGYIFDDTKDINEIKLGQNA
 FIFAVCEYLKHNPNKQYLCVAQKVARGILSMINQDTYETHILNYPDLTVKESFRIIYD
 GEAAALALLRLYHQDNDKWLEVVKKLMDRFIEKEYWQYHDHWLGYCTNELVQLCPQDKYF
 EFGIKNVNTYLEYIEQRETTFPTFLEMLMATYKLIQKAKATHRQELVTQLIDEKLINVI

HTRAELYQVRGFFYPEIAMYFKNPKRILGSFFIKHHGYRVRIDDIEHYISGYVQYQKAQIK
DEIL*

Sequence 39

5 Contig_0442_pos_5645_5175,
putative peptide of unknown function
atgttttagttatcaaataaaataaaatattaaattaaaaatattagaagaacgagaagcc
gaacagttatattaaattagtagatagcaatcgtgactatttagctgaatttctgcctttt
gttgaaacatacgaagaaagttgaagatagtaaacactttatccattcggcggttgcaacaa
10 tttatcgatggtaatggatttcattgtggaatatggagtaataaagaattgattggagtc
ataggattgcattacttagatttagttaataaaacaacttcaattggttattatttagct
gaagactttcaaaagaaaggtattatgactaaatgtactaaagcgtaattcgcctatgta
tatgaagtgtatgatattaatcgtatagaaatacgaatgtctactaaaaataagaaaagc
agagctataccaattagacttgggttcacgcaaaggtatattgagaagtaa

Sequence 40

MFSYQINKNIKLKILEEREAEQLFKLVDSNRDYLAEFLPFVEHTKKVEDSKHFIHSALQQ
FIDNGFHCIGIWSNKELIGVIGLHYLDLVNKTTSIGYYLAEDFQKKGIMTKTKALIRYV
20 YEVDINRIEIRMSTKNKKSRAIPIRLGFTQRYIEK*

Sequence 41

Contig_0442_pos_0_925,
is similar to (with p-value 3.0e-20)
>pir:pir|S52351|S52351 hypothetical protein 1 - Staphylococ-
25 us xylosus >gp:gp|X84332|SXGKG2_1 S.xylosus glucose kinase g
ene. NID: g666114.
atgctaattaataatgaagataaaaggacttaccttcattacaaaagaaaagttttaaca
caaaatcttggtgataaacatatgcaacgttttaccctattacatatacacttatctta
attaatattgtgatattggttatgtatgattttatacttaaatcgattttctgatgttaaa
30 ctattagaagtaggtggaacttggttcattttaatggtgttcacggagaatggtatagactt
atttcgtcaatgtttttacatttttaatttcgaacacattttaatgaatatgctctctcta
tttatttttggttaaaattgtcgaatcaatcattggatcatggcgaatgctaataatttat
ataatatccgattatatggaaattttgtttctctatcatttaatacgactacaatttca
gtcgggtgctagtggagcaatatttgggtctaattgggttctattttgtgattatgtattta
35 agcaagaattttaataaaaaaatgattggccagttattaattgcttttggttggttttaac
gttttttcactttttatgtctaatattaatataatggcacatttaggtggatttatcagt
ggtgtattaattacattaataggctattattttcaaacacacagctctttattttggtca
tttttgattgtatttttacttatattcatcattttacaaattagaatatttactataagt
gaggataatatctatgataaatttaattcgggatgaaatgattaaggtaattatagcgaa
40 gcaaaaaatggttgtaaaacaaacacttaataataattacgccgatgatgaaacatattac
cttagtggtttgattactgcaactaagagttcgcaagcagaggccgatcagaatgggaa
agaggtttaagaaaatttccaaatt

Sequence 42

45 MLINNEKRTYLYHYKRKVLTONLVDKHMQRFTPTITYTLILINIVWLICMILYNRFSQVK
LLEVGGGLVHFNVVHGEWYRLISSMFLHFNFEHILMNMLSLFIFGKIVESIGSWRMLIY
IISGLYGNFVLSFNNTTISVGASGAIFGLIGSIFVIMYLSKNFNKMKMIGQLLIALVLI
VFSLFMSNINIMAHLGGFISGVLITLIGYYFKTQRSFLWSFLIVLLIFILQIRIFTIS
EDNIYDKLIRDEMIKGNYSKNNVVKQTLNNNYADDETYLISGLITATKSSQAEAVSEWE
50 RGLRKFPNX

Sequence 43

Contig_0443_pos_545_1741,
putative peptide of unknown function
55 atgggttaaattttatacactgtgctgatttgacatttgacagtcctttcaaattctaaaagt
tatcttagtcctaaatattttgaagatgtccaaaagagtgcatatgaaagttttaaaac
atagtcgacttagctttaaacaggaagtcgattttattattatagcaggtgatttattt
gatagtgagaatcgtacattgcgtgctgaagtcttttttaaatagaacaatttgaaagatta
agaaaaagaacaaatatttggtttatatttgccatggcaaccacgatcctcttacttctaaa

ataacaagtcagtgccctaataacgtatccgtatTTTTCAAATCAAGTAGAGACATATCAA
 gctatcactaaatcaggagaaacaatttatattcatggattcagctatcaaaatgatgcg
 agttatgaaaataaaatagacgcatacccatcaagtcaggtcagaagggcatacatatt
 ggtgtattacatggaacttatagtaaatcttcggtgaaagaccgttatactgaatttagg
 5 ttagaagacttaaatcaacgtttataccactactgggcattaggacatatacaccaacgt
 gaacagttaagtgcacatgccagtcattaactatccaggtaatatccaaggaagacatttc
 aatgaattaggagaaaaagggttgtctattggcggaaggtgatcatcttaaactcactaca
 caattttatcctactcaattttattaaatttgaagaagctacaattgaaactgatcataca
 tctaaacaaggactttatgatgttattcaatcttttaaagataaagtaagaactgaaggg
 10 aaatcattttatagattgaacgtacgcattaatagtgaagacattattgcaccacaagat
 ttaattcaattaaaagaaatgattactgagttcgaagaaaacgaaaatcaatttggtttt
 attgaagatttaaatcttcaatatgttcaaaatgacgaaatgccaatagttaaagagttt
 tcaccagaattacttgatgatgcgtcactttttgattcggcaatgactgatttatatctt
 aatccaagggtcttcaagtttttagatgactataatgaatttgataaagttgagtttagtc
 15 aatcatgcagaaagacttttaaaggatgaaatgagaggtgaacaaaatgataattaa

Sequence 44

MVKFIHCADLHLDSPFKSKSYLSPNIFEDVQKSAYESFKNIVDLALKQEVDFIIAGDLF
 DSENRTLRAEVFLNEQFERLRKEQIFVYICHGNHDPLTSKITSQWPNNVSVFSNQVETYQ
 20 AITKSGETIYIHGFSYQNDASYENKIDAYPSSQGQKGIHIGVLHGTYSKSSVKDRYTEFR
 LEDLNQRLYHYWALGHIHQREQLSDMPVINYPGNIQGRHFNELGEKGCLLVEGDHLKLT
 QFYPTQFIKFEEATIETDHTSKQGLYDVIQSFKDKVRTEGKSFYRLNVRINSEDIAPQD
 LIQLKEMITEFEENENQFVFIEDLNLQYVQNDempiVKEFSPELLDDASLFDSAMTDLYL
 NPRASKFLDDYNEFDKVELVNHAERLLKDEMARGEQNDN*
 25

Sequence 45

Contig_0443_pos_2433_0,
 is similar to (with p-value 1.0e-45)
 >sp:sp|P54596|YHCL_BACSU HYPOTHETICAL 49.0 KD PROTEIN IN CSP
 30 B-GLPP INTERGENIC REGION. >gp:gp|X96983|BS75DGREG_13 B.subtilis
 chromosomal DNA (region 75 degrees: cspB upstream of glp
 PFKD operon). NID: g1239975. >gp:gp|Z99108|BSUB0005_181 Bacillus
 subtilis complete genome (section 5 of 21): from 802821
 to 1011250. NID: g2633055.
 35 atgcatgagcaaaaacaaaagaggttgctctacacgatcaaacacaagaatggaaaagg
 ttagaacagtcgcttaatatagagcctataaattttcctgaaaaagggatagatagatac
 gaaactgctaaatctcacaacaatcacttgaacgagataaaaagtttgcgagaagaaaga
 ttaagcatattaaataaagaggcggagtcctcatcaatccagtagaccaaaagtatatgtat
 tcgtttaatagcctttatcaacaagagactgaaattaaacaaaaagaatttgagttacgt
 40 tcaatagagaaagatatgtctgataagcaacgtgaactagaagctcttcaatctataggt
 atcgattttggcattgtattacacctcatatatggtgcagagctctaaaactctcgaacaa
 tcaacagactggtttagtagtattgttgagatggttatgttgcaactattacaaatgattgtc
 atgccactaatattcatttcaattggtgccgcttttagcaaaaatacaaatggtgaaaaa
 ttcgctaagatcggttcttatattttatggttttaattggtactgtagccattgcagct
 45 atcgttggaattttttacgctttgatctttggttttagatgcatcgctctattgatttaggt
 agtgcagaacattcacgtggtacagaaatttcaaaaacaagccaaagatttaactgcaaac
 actttaccacaacaaattctcgaagtattcccaagcaatccatttttagatttcacagga
 caacgtacaacttcgacaattgcagttggtatttttgcaacggttggtgggctttgcttat
 cttagagttgcaagaaaacagccggaacatggaagcttactt
 50

Sequence 46

MHEQKQKEVALHDQTEWKRLEQSLNIEPINFPEKIDRYETAKSHKQSLERDKSLREER
 LSILNKEAESINPVDQKYIDSFNSLYQQETEIKQKEFELRSIEKDIADKQRELEALQSIG
 IVFGIVLHLIYGAESKTLEQSTDWFSIVGDGYVALLQIMVPLIFISIVAAFSKIQIGEK
 55 FAKIGSYIFMFLIGTVAIAAIVGIFYALIFGLDASSIDLGSAEHSRGTEISKQAKDLTAN
 TLPQQILEVFPSPNFLDFTGQRTTSTIAVVI FAFVGFAYLRVARKQPEHGSLL

Sequence 47

Contig_0444_pos_4472_4089,

is similar to (with p-value 7.0e-18)

>gp:gp|U40604|LMU40604_2 *Listeria monocytogenes* ClpC ATPase (mec) gene, complete cds. NID: g1314293.

5 gtgaaacagatacttcaacaccttgctgcaaaacatggtattaattttcatgagatggca
tttaaagaagaaaaaaatgcccaacgtgtcagatgacacttaaggatattgcacatggt
ggtaaagcttgggtgtgctgattgttatgctacgtttaaagaagacatcattgatatagtt
caacgtgttcaaggtggtcaatttgaacatgtaggaaaaacaccacaatcatcgatataag
aaacttgcaataaaaaagcaaattgaagaaaaatcaaaatatctaaataaattgatagat
10 ggtcaagagtttgaagaggcagcgattgttcgtgatgaaattaaagctttaaaagtgag
agcgaggtgtctcatgatgagtaa

Sequence 48

VKQILQHLAAKHGINFHEMAFKEEKKPTCQMTLKDIAHVGKLGACDCYATFKEDIIDIV
QRVQGGQFEHVGTKPQSSYKKLAIKKQIEEKSKYLNKLIDGQEFEEAAIVRDEIKALKSE
15 SEVSHDE*

Sequence 49

Contig_0444_pos_3078_625,

is similar to (with p-value 0.0e+00)

20 >sp:sp|P37571|MECB_BACSU NEGATIVE REGULATOR OF GENETIC COMPE
TENCE MECB. >gp:gp|D26185|BAC180K_148 *B. subtilis* DNA, 180 k
ilobase region of replication origin. NID: g467326. >gp:gp|U
02604|BSU02604_2 *Bacillus subtilis* Marburg 168 ClpC adenosin
e triphosphatase (mecB) gene, complete cds, orfX and orfY, p
25 artial cds. NID: g442358. >gp:gp|Z99104|BSUB0001_86 *Bacillus*
subtilis complete genome (section 1 of 21): from 1 to 21308
0. NID: g2632267.

atgttatTTGGtagattgacagagcgtgcacaacgtgtgttggcacatgcacaagaggaa
gcaattcgTTTgaaccattctaatttgaacagaacatctTTTgcttggTTtaatgaaa
30 gagccagaaggtatagcagcaaaaggtattagtaagTTtaattactgaagataaagtc
atcgaagaagTTgaaaaacttatcggtcacggTcaagagcaaatgggcacactacattat
acaccgagagcaaaaaagtaattgaactgtctatggatgaagctcgaaagctacatcat
aactTTTtaggaacagagcatatactattaggtTTaattagagaaaaatgaaggtgTTgca
gcacgtgttatTTgcaaaccttagattTTaataattactaaagcacgtgcccaagTTgtaaaa
35 gctTTtaggaagTccagaaatgagtaataaaaaatgcgcaagctaataagtctaataacacg
cctactTTtagatggattagctagagattTaaactgttattgctaaagatggaacgTTtagat
ccagtcgtaggacgagataaagaaattactcgtgtaattgaagTTTtaagtcgTcgTact
aaaaataatcctgtgctaattggtgaacccggTgTTggtaaaaacagcaattgctgaaggg
cttgcgcaagcaattgTtaaaaaatgaagtaccagaaaactTTaaaagacaaacgtgtaatg
40 tcattagatatgggtacagtcgtagctggcactaaatatcgtggTgaattgaaagaaaga
TTgaaaaagTTtagggagaaatccatcaagctggtaatgTTattctatttatcgatgaa
cttcatactTTtagttggcgtggtggcgagaggagcaattgatgcattctaattTTta
aaacctgctTTtagctcgTggagaattgcaatgtatagtgccacaacattagatgaatat
cgtaaaaaatatagaaaaagacgctgcattagaacgtcgtTTTcaaccaattcaagtggt
45 gaacctacagTTgaagacacgattgaaatctTaaaaggattacgtgaccgTTtatgaggct
catcacagaattaatatctcagatgaagctTTtagaagcggctgctaaattgagtgatcgc
tatgTTTcagatcgTTTctTgccagataaagccattgactTaatTgatgaggcaagTTca
aaagTTtagactTaaaagtcatacaacgccaagtaattTaaaagagattgaacaagaaatt
gataaagTaaaaaatgaaaaagatgctgcagTTcatgctcaagaattTgaaaatgccgct
50 aattTaaagataagcaatctaaactTgaaaagcaatatgaagatgctaaaaatgaatgg
aaaaatgcacaaggtggtTTtagatactgccttatctgaagaaaatatcgctgaagtaata
gctggtTggacaggtattcctTTtaactaaaattaatgaaactgaatcagatcgTTtattg
aatctTgaagatacactTcataaacgtgtcattTggacaaaacgatgctgtcaattcaatt
agtTaaagctgTTtagaagagctcgTggtgTctTaaagatccaaaacgtccaatcggtagt
55 tttattTTcttaggacctacaggtgTgggtTaaaactgaattggctcgTgctTTtagctgaa
tctatgTTTggtgaagacgatgcaatgattcgctgtagatatgagtgaaTTtatggagaaa
catgctgtcagtcgattgTTggtgcacctccaggatatgtaggacatgatgacggcggt
caattgactgaaaaagTtagacgTaaaccatactctgtgattTTTattTgatgaaattgag
aaagcacatcctgacgtattTaatattctTctacaagTTTtagatgatggtcattTaaaca

gataactaaaggctcgactgtggacttccgtaatactgtgattattatgacttctaagtgtg
 ggagctcaagaattacaggaccaacgctttgctgggtttggaggtgcttcagaaggtagt
 gactacgaaactgtcagaaaaacaatgatgaaagaattaaaaaattcattccgaccagaa
 5 ttccttaaaccgtgttgatgacattattgtcttccacaaacttacaaaagatgaattaaaa
 gaaattgttacaatgatggtaataaaacttactcaccgtctttcagagcaaaatattaat
 attgttggttactgataaagcgaagaaaaaattgcagaagaaggatatgatcctgaatat
 ggtgctagaccactcattagagcaattcaaaaaacggttgaagataatttaagcgaattg
 attttagatggaataaaaattgaaggtaaagaagtaacaattgatcatgatggtaagaa
 10 ttttaagtatgatatttatgaaattacagctaaaaaagaacaacagaatcataa

Sequence 50

MLFGRLTERAQRVLHAHQEEAIRLNHNSNIGTEHLLLGLMKEPEGIAAKVLVSFNITEDKV
 IEEVEKLIGHGQEQMTLHYTPRAKKVIELSMDEARKLHHNFVGTEHILLGLIRENEGVA
 ARVFANLDLNLTKARAQVVKALGSPMSNKNQAANKSNNTPTLDGLARDLTVIAKDGTLN
 15 PVVGRDKEITRIVIEVLSRRTKNNPVLIGEPGVGKTAIEGLAQAIKNEVPETLKDKRVM
 SLDMGTVVAGTKYRGEFEERLKKVMEEIHQAGNVILFIDELHTLVGAGGAEGDAIDASNIL
 KPALARGELQCIGATTLDYRKNIEKDAALERRFQPIQVDEPTVEDTIEILKGLRDRYEA
 HHRINISDEALEAAAKLSDRYVSDRFLPDKAIDLIDEASSKVRLKSHTTPSNLKEIEQEI
 DKVKNKDAAVHAQEFENANLDRKQSKLEKQYEDAKNEWKNAQGGLOTALSEENIAEVI
 20 AGWTGIPLTKINETESDRLLNLEDTLHKRVIGQNDVNSISKAVRRARAGLKDPRPIGS
 FIFLGPTGVGKTELARALAESMFGEEDAMIRVDMSEFMKHAVSRLVGAPPGYVGHDDGG
 QLTEKVRKPYSVILFDEIEKAHPDVFNILLQVLDDGHLTDTKGRTVDFRNTVIIMTSNV
 GAQELQDQRFAGFGGASEGSDYETVRKTMKELKNSFRPEFLNRVDDIIVFHKLTKDELK
 EIVTMMVNKLTHRLSEQNINIVVTDKAKEKIAEEGYDPEYGARPLIRAIQKTVEDNLSEL
 25 ILDGNKIEGKEVTIDHDGKEFKYDIYEITAKKETTES*

Sequence 51

Contig_0445_pos_1513_1908,
 is similar to (with p-value 1.0e-34)
 30 >gp:gp|AF009352|AF009352_4 Bacillus subtilis osmoprotectant
 transport system OpuC including ATPase (opuCA), transmembran
 e protein (opuCB), osmoprotectant binding protein precursor
 (opuCC) and transmembrane protein (opuCD) genes, complete cd
 s. NID: g2271388.
 35 atgattgaatgtaacaaattacctttcattacttatgaccacctttcttttcttcaaagt
 aatgatgtttctttaaatattcttcagctatcactgcaggttcttaccttttccatccg
 cttcataatttaacttctgcatttcttctgttgagattttaccttctaatttttttagtg
 ccttatcgatttctggattatcctttatattaattgttcatattgcaagtggactaccgtcat
 aaggcgggaagaatttgcgatcatcttccaatattttcaaatacataagctgcaatacgtc
 40 catctgttgtaatacccaactgtacatctaattttatttttttaatgcatcatatacta
 aaccaatttgcatggacgtgcactatcaaattta

Sequence 52

MIECNKLPFITYDHLQSLQNDVSLNQLSLQVPYLFHPLHNLTSAFLLLRFYLLIFLV
 45 PYRFLDYPLLIQVDYRHKAGRICDHLPIFSNHKLQYVHLLNTQLLHLIYYFLMHHL
 NQFALDVHYQI*

Sequence 53

Contig_0445_pos_8150_8581,
 putative peptide of unknown function
 50 atgttcaaaaatatattattaccctatgattttcgaaaatgatttttagtgctatccctgac
 tatttagaaaaagtcaccgatgaagattcagttgttgtaatttatcacgttgtaacagaa
 aatgatcttgcaatttagtgtcaagattataataagcataaagaagatattattagagaa
 aaagagaaaaaaactcactcatttttacgtgaattagaaaaaagagatattcaatataaa
 55 atagatgtagattttgggcatattaaagatacaatcttagaaaaaattacttctggagat
 ataaataatgggtgaatttgatttagtaattatgagtaatcatagagtcgatttgatatt
 aaacatgttttaggagatgttacacataagattgctaaaagaagttctgtcccagtacta
 attgttaataaa

Sequence 54

MFKNILLPYDFENDFS AIPDYLEKVTD EDSVVVIYHVVTENDLAISVKYYNKHKEDI IRE
KEKKLTPFLRELEKRD IQYKIDVDFGH IKDTILEKITSGDINNGEFDLVIMSNHRVDLNI
KHVLGDVTHKIAKRSSVPVLIVK*

5

Sequence 55

Contig_0445_pos_7486_7115,

10 putative peptide of unknown function

gtggattacatggcaagacattttggagtttactatagcttgacaactatttctcgtgac
ttacaagaatttagaaatttacaaaatccctgttgaaaataaaaagtatatattacaagaaa
ataaatcaaacaaatcaattaagtgcacaaaaacaattagaaatatttagtgatgagatt
attgaatttataacgctaaataactatgtcttaataaaaacatctcctggccttgctcaa
15 agtataagttattacatagatcaattacaaatgaaagaaatattaggaattattggaggt
aacgatactttgatgattttgacttcttcaaatgaaatagcagaatttggttattatcaa
ttattcccttaa

Sequence 56

20 VDYMARHFGVYYSLTTISRDLQLEIYKIPVENKKYIYKKINQTNQLSAKKQLEIFSDEI
IEFITLNNYVLIKTSPGFAQSISYYIDQLQMKELGIIGNDTLMILTSSNEIAEFVICYQ
LFP*

Sequence 57

25 Contig_0445_pos_6795_5611,

is similar to (with p-value 0.0e+00)

>gp:gp|Y17554|BLY17554_1 Bacillus licheniformis arcA, arcB,
arcC and arcD genes. NID: q3687415.

gtgttgtaaaaagaccaggaagaattagaaaatttagtacctgatcatttaagtggg
30 ttattattcgatgatattccctacttaaaagttgcacaaagagcatgacaaatttgct
caaactttgagagatgaaggaatcgaagtagtttatttagaaaaacttgcagcagaatct
attactgagccagaagtacgcgagaacttcataaacgacatattaacagaatctaaaaag
acaatttaggtcatgaaactgaaattaaagaattcttttcaaagttatctgaccaagaa
cttgtaataaaatcatggctggcgtacgtaaagaagaattcaacttgaaacaacccat
35 ttagtagaatatatggatgatagatatccattttacttagatccaatgcccaacctttat
tttacaagagatcccccaagcttcaattggttagaggaatgacaattaacagaatgtattgg
agagcacgacgtagagaatctatttttatgacatatatactgaaacatcatccaagattt
aaagataaagatgtaccagtatggttagatcgtaactcaccatttaattgaaggtgga
gatgaattagattattcgaaagatgttttagctattggtatatcagaacgtacatcagct
40 caagcaatagaaaagttagcacgtaatattttcaaagatgcaaacacaagttttaaaaa
atcgtagctattgaaatacctaatacacgtacattttatgcacctagatacagtactaact
atgattgactacgataagtttacagtacatgcagcaatatttaagaagaaaaataatg
aatatatttaccatagaacaaaatgatggttaaggacgatataaaaattactcgttctagc
aagttacgtgaaacacttgctgaagttttagaagtagaaaaagtggaactttattccaaca
45 ggtaatggcgacgttattgatggtgcacgtgaacaatggaatgatggctcaaacacatta
tgtattcgaccaggggttggtgacatacgcacgtcgcaactatgtatcaaaccaactttta
cgcgacaaaggaattaaagtgattgaaattactggtagtgaacttgtagctggacgcgga
ggcccaagatgtatgagtcagccgttatttagagaagatattta

Sequence 58

50 VLLKRPKGELLENLVPDHL SGLLFDDIPYLKVAQEEHDKFAQTLRDEGIEVVYLEKLAES
ITEFEVRENFINDI LTESKKTILGHETEIKEFFSKLSDQELVNKIMAGVRKEEIQLETH
LVEYMDDRYPFYLDPMNLYFTRDPQASIGRGMTINRMYWRARRRRESIFMTYILKHHPRF
KDKDVPVWLDNRNSPFNIEGGDELVLSKOVLAIGISERTSAQAIEKLARNIFKDANTSFKK
55 IVAIEIPNTRTFMHLDTVLTMDYDKFTVHAAIFKEENNMNIFTIEQNDGKDDIKITRSS
KLRETLAEVLEVEKVDFTPTGNGDVIDGAREQWNDGSNTLCIRPGVVVITYDRNYVSNQLL
RDKGIKVIEITGSELVRGRGGPRCMSGPLFREDI*

Sequence 59

Contig_0445_pos_5525_4104,

is similar to (with p-value 0.0e+00)

5 >gp:gp|Y17554|BLY17554_3 Bacillus licheniformis arcA, arcB,
arcC and arcD genes. NID: g3687415.
atggatgaaaataaattaggtaaaacttccttaattgggttagtcataggctctatgata
ggcgggtggtgcattcaatatcatctcagatatgggtggccaagctgggtgacttgcaata
attatcgggttgataataactgctattgggtatgatttctcttgcgtttcgtatttcaaaat
10 ttaacaaatgagcgaccagatcttgatggaggaatttatagttatgctcaaacagggttt
ggagatttttattgggttttcaagtgcctgggatattgggttgcagcatttctaggaat
gtggttatgcaaccctattaatgtcagctgtgggtaacttttccctatatttaaagga
ggtaacacacttccaagtattatcatagcatcaattttattatggggtgtacatttttta
atacttagaggtgtagaaactgcagcgtttataaatagttattgtacagtagctaaatta
15 ataccatatatttctagttatttatatgcatgatagttgtattcaacttcagtacttttaaa
tccggtttttatgggtatgactagtggaagtgttggcgtttttagttggggagatacaatg
gcacaagtaaaaagtactatgttagtaactgtatgggtattcacagggttgaaggagcc
gtgtcttttctctggacgtgcaaaagtctaaaaaggtatgtaggaactgctaccgttattggg
ttgatttctgtgctagtcattttatttcttaagtactgtactagcccaaggtgtcattcag
20 cagaaccaaatttcaaaaacttgctaattccatcaatggcacaagtattagaacatattgta
ggtcattgggggttcagtgttagttaatataggcttaattatctctgttttaggagcttgg
ttaggatggacattactagctgggtgaattaccattcattgttagctaaagatggacttttc
ccgaaatgggttgctaagaaaaataagaataaagctccggtcaacgctttaattattact
aatatattagttcagttatttttaattagttatgttggttacagatagtgccatcagttt
25 gcgttttccacttgcatcaagtgcattcttaattccatatatactcagtgctttttaccag
gttaaatatactattcaaaaataaatctaaagctaattttaaacaatggataataggaatt
attgcatctatttacacaatttggttggtttatgcagctggattagattatttactatta
acgatgttggttatatatacctggattactcgtatacagctacgtacaaagggataalaac
aaacatttgacaaaattggattatacgttattcatattcatcattgtacttgcaataata
30 ggaatagttcgtttgattacaggtaatatttctgtattttaa

Sequence 60

MDENKLGKTSLIGLVIGSMIGGGAFNIISDMGGQAGGLAIIGWIITAIGMISLAFVFQN
LTNERPDLGGIYSYAQTGFDFIGFSSAWGYWFAAFLGNVAYATLLMSAVGNFFPIFKG
35 GNTLPSIIIIASILLWGVHFLILRGVETAAFINISIVTVAKLPIIFLVIICMIVVENFSTFK
SGFYGMTSGSVGVFSWGDMAQVKSTMLVTWVFTGIEGAVVFSGRAKSKKDVGTATVIG
LISVLVIYFLMTVLAQGVIOQNQISKLANPSMAQVLEHIVGHWSVLVNIGLIISVLGAW
LGWTLLAGELPFIVAKDGLFPKWFAKENKNKAPVNALIITNILVQLFLISMLFTDSAYQF
AFSLASSAILIPYTLFAFYQVKYTIQNKSKANLKQWIIGIIASIYTIWLVAAGLDYLLL
40 TMLLYIPGLLVYSYVQRDNNKHLTKLDYTLFIFIIVLAIIGIVRLITGNISVF*

Sequence 61

Contig_0445_pos_4062_3373,

putative peptide of unknown function

45 atgtatgaagaaaatatttatattaaaaattcagaatatgaatttgataataatctttaa
caattagcatcatacttaaatattcctgttagtattggttagacctataaaaggagattta
acactttatcaatataaaaaaggacaagtcatatatcattcaactgatcaataaaattt
gtatactttttagtaaatggttgattttacatgaatcttctaataattactggtgacaat
tatttaagattaagtaaaagacgaaaatataatttccaatgaacttcataatttaataaacc
50 cctgcaccatatgaaatatgtacagctttgacagattgtaaaaatattaactttaccgaaa
gatttacttgagtatttatgtagaaagcataatgaaatatttgaaagtctcttcaagaaa
cttaatgagactattcaatttcaagtagaataatattatggcgttaagagctaattcagct
aaagaagaattgaaagaatactacaaaattttatgcctttcaattggggatgalaatgga
gaattctatgaattaaaacaaattatgactgttcaattaataagtaatttatctggactt
55 aacagaaaaactactggtgaaataatcagagaattaaaaatagaaaatattatatatcaa
gataaaagaatttgattataaaaaataa

Sequence 62

MYEENIYIKNSEYEFDNNLKQLASYLNIPVSIVRPYKEDLTLYQYKKGQVIYHSTDQIKF
VYFLVNGCILHESSNITGDNYLRLSKDENIFPMNFIFNETPAPYEICTALTDCIKILTLPK
DLLEYLCRKHNEIFESLFKKLNETIQFQVEYIMALRANSAKERIERILQILCLSIGDDNG
5 EFYELKQIMTVQLISNLSGLNRKTTGEIIRELKIEENIYQDKRNWIIK*

Sequence 63

Contig_0445_pos_3371_2502,
is similar to (with p-value 7.0e-53)
10 >gp:gp|AF009352|AF009352_3 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
(opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_68 Bacillus subtili
15 s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: g2635827.
atgaataaaatattaattgaaaaggagatattcaaaatgaaaatttaagaaatagaac
tttttaactttgttagacttcacacaaaagaatggaatttttacttaatttatctgaa
gatcttaaacgcgcgcaaaatatgcaggaatagaacaacaaaaaatgaaaggtaaaaatatac
20 gctctactttttgaaaagattcaacacgcactcgatgtgcatttgaaacagcggcttat
gatcaagggtgcacatgtaacataccttgggccaacaggttctcaaatgggtaaaaaagag
tctaccaaagatactgctcgtgttttaggtggagctgtcccttaggtattttattatca
aaaacgcaacgcacagctaattgtggtattaacagttgctggcgtgcttcaaaccattcct
actttggctgtgctagctatcatgattccaatatltggggtaggaaaaacaccagctatt
25 gttgcattatttatctatgtattattaccaatttttaataataacagtattaggtgttaa
aatatcgataaaaaatgtcattcaagctgggtcaaagtatgggaatgactaaatttcaatta
atgaaagatgtagaaatgccttttagctttaccacttattattagtggtattcgtctatca
agtgtatacgtcattagttgggcaacactcgcaagttatgtaggtgcaggtggacttggg
gatcttgattttaattggattaaatctctatcaaccacctatgattattagtgacagcatt
30 gttgttactttattagcattagttattgactttatactttcattagttgaaaaatgggtt
gtacctaaaggattaaaagatctagataa

Sequence 64

MNKILIEKEIFKMKNLRNRNFLTLLDFTQKEMEFLNLSEDLKRAKYAGIEQQMKMGKNI
35 ALLFEKDSTRTRCAPETAAYDQGAHVTYLGPTGSQMGKKESTKDTARVLGGAVPLGILLS
KTQRTANVVLTVAGVLQTIPTLAVLAIMIPIFGVGKTPAIVALFIYVLLPILNNTVLGVK
NIDKNVIQAGQSMGMTKFQLMKDVEMPLALPLIISGIRLSSVYVISWATLASYVGAGGLG
DLVFNGNLNLYQPPMIISAAIVVTLALVIDFILSLVEKWVVPKGLKVS*

Sequence 65

Contig_0445_pos_2434_1541,
is similar to (with p-value 7.0e-85)
40 >gp:gp|AF009352|AF009352_4 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
45 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388.
gtgttatctggatgcagtttaccaggtttaggtgatgaaatgcaaaagatgatgtgaaa
atcacacgactgaaacaagtgaactaagattataggtcatatggaaaaattattaatt
50 gaacatgaaactgatggaaaaatcaaacgcgcttgattgggaacctaggttctagcatt
attcaacataatgcgttacaacgtggtgatgcaaatatgtcagcgggtacgttacacaggt
actgaattgacaggtgtattagcagctaaacctactaaagatcctgataaggccatgtct
gaaacacacgcttatttaaaaagaatatgatgaaaagtattatcattcacttgggttt
gcgaatacatacgcattcatggtgacaaaagaaacggctaaaaagtatcacttagaaaaa
55 gtatcagatttagagaaatataaagatgaactacgtcttggaatggatacccaatggatg
aaccgtgcaggtgatggatatccagcttttgtaaagattatggatttaaatttgatagt
gcacgtccaatgcaaatgggttttagtatatgatgcattaaaaataataaattagatgta
gcagttgggtattcaacagatggacgtattgcagcttatgatttgaaaaatttggaagat
gatcgcaaatcttccgccttatgacggtagtccacttgcaaatgaacaattaataaag

gataatccagaaatcgataaggcactaaaaaatttagaaggtaaaatctcaacagaaagaa
atgcagaagttaaattatgaagcggatggaaaaggttaaggaacctgcagtgatagctgaa
gaatattttaagaacatcattactttgaagaaaagaagggtggtcataagtaa

5 Sequence 66

VLSGCSLPGLGDGNAKDDVKITTTTSETKIIHMEKLLIEHETDGKIKPTLIGNLGSSI
IQHNALRGDANMSAVRYTGTELTSVLAAPKTPDKPDKAMSETQRLFKKKYDEKYHSLGF
ANTYAFMVTKETAKKHYHLEKVSLEKYKDELRLGMDTQWMNRAGDGYPAFVKDYGFKDS
ARPMQIGLVYDALKNNKLDVAVGYS TDGRIAAYPDKILEDDRKFPPYDGSPLANEQLIK
DNPEIDKALKKLEGI STEEMQKLN YEADGKGKEPAVIAEYLLKHHYFEEKKGHHK*

Sequence 67

Contig_0445_pos_1454_846,
is similar to (with p-value 9.0e-30)

15 >gp:gp|AF009352|AF009352_5 *Bacillus subtilis* osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
(opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_66 *Bacillus subtili*
20 s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: q2635827.

atgtcggtatatggtgtgttgtttgcatgtataattggaattcctattggtattttcata
gccaagtataaacgtttatcgtggccggttaattacaattgcaaattatatacaaaactgtt
ccagcaatcgctatgttagccatacttatgttggctatgggattaggaccaacaactgtt
25 gttgtaactgtattcctatatctcgttattacctattattaaaaatacttatactggtatt
gtagaagttgatgaaaatattaagacgctggtaaaggatatgggaatgacggggaatcaa
atattaagaatgatagagttaccattatctttatctgttattattgggtggtgttagaatt
gcacttgttgttgcctatcggaatagtagcgattgggtcatttatcgggtgctccaacacta
ggtgatattattattcgtggtacaaattcaacagatggaacaacattcatcttagcaggt
30 gccataccaattgctttaatagcaattatcatagatataggattacgttatctagaaaaa
cgtttagatcctactcgtaaaaacaaaaagattcaatgcaaaaacatcaagtacaaaaa
ttacqtaa

Sequence 68

35 MSVYGVLFACIIIGIPIGIFI AKYKRLSWPVITIANIIQTVP AIAMLAI LMLAMGLGPTTV
VVTVFLYSLLP IIKNTYTGIVEVDENIKDAGKGMGMTGNQILRMIELPLSLSVIIGGVRI
ALVVAIGIVAIGSF IGAPTLGDIIRGTNSTDGTTFILAGAIPIALIAIIIDIGLRYLEK
RLDPTRKNKKDSMOKHQVOKLR*

40 Sequence 69

Contig_0446_pos_520_1677,
is similar to (with p-value 4.0e-36)

45 .>gp:gp|AF008930|AF008930_4 *Bacillus subtilis* choline transpo
rt system including ATPase (opuBA), transmembrane protein (o
puBB), choline binding protein precursor (opuBC) and transme
mbrane protein (opuBD) genes, complete cds; and unknown gene
. NID: q3068551. >gp:gp|Z99121|BSUB0018_57 *Bacillus subtilis*
complete genome (section 18 of 21): from 3399551 to 3609060
. NID: q2635827.

50 atgaaaccacttagaagattgaccaaagtcgaactccctattgcaatgctgttatcatg
gcaggaatacgcacagctatgggtattaatcattgggtactgctacactcgcagctttaata
ggcgctggtggtctaggagatttaatatattattaggcattgatcgtaacaatagtgcactc
attttaatagggtctattccagctgcacttctagctattatTTTTGATTttattttaaga
55 tacatggaacgtttatcatataaaaaattgctcatttctttagggacaattgtaatttg
attatcatagctattgccatacctatggcagcgcaaaaagggtgataaaatcacattcgca
ggcaagctagggttcagaaccgtcaattattacgaatatgtataaaatacttattgaagaa
gacacagatgatactgtagaagtcaaagatggcattgggtaaaacctctattttaat
tgcgttaagtgcagatgaaattgatggttatttgaatttacaggtactgtattaggtgaa
ttaacgaaagaagatttaaagtcataaaaaagaaaacgatgtatatcaacaagcaaaagtct

agtttagaaaaaaatatgatatgacaatgcttaaacgatgaaatataataacgtat
gcattagctgtataaacgtgactttgcaaaaaaatatcaaattagacaataggtgattta
cgcaaggtagaagataaaacttaaaccaggttttacattggaatttaatatgacaccagat
ggatacaaagctgttaaaaaaacgtatcatcttaatctttctaattgttaaaactatggaa
5 cctaaattacgtttatactgcagttaaaaaggagatattaatctcatagacgcatactct
actgatgcagaattaaaaacaatataacatggtagtattaaaagatgatcaacatgtattt
cctccataccaaggagcaccgctattttaagaaaaaatatttaaaagaccatcctgaagtt
aaaaaacgctcaataaattggcgaatagaatcacagatgaagaaatgcaagaaatgaac
10 tataaggtaacagtgaagaaaggatccttataaagtagcaagagaatacttagaaaaa
gaaaaattaataaaataa

Sequence 70

MKPLRRLTKVELPIAMPVIMAGIRTAMVLIIGTATLAALIGAGGLGDLILLGIDRNSAL
ILIGAIPAALLAIIFFDILRYMERLSYKKLLISLGTIVIVIIIAIAPMAAQKGDKITFA
15 GKLGSEPSIITNMYKILIEEDTDDTVEVKDGMGKTSFLFNALKSDEIDGYLEFTGTVLGE
LTKEIDLKSKKENDVYQAKSSLEKKYDMTMLKPMKYNNYALAVKRDFAKKYQIKTIGDL
RKVEDKLPKPGFTLEFNDRPDGYKAVKKTYHLNLSNVKTMPEKLRYTAVKKGDINLIDAYS
TDAELKQYNMVVLKDDQHVFPYQGAFLFKEKYLKDHPEVKKPLNKLANRITDEEMQEMN
YKVTVKKEDPYKVAREYLEKEKLIK*

Sequence 71

Contig_0446_pos_3200_2466,
is similar to (with p-value 3.0e-77)
>sp:sp|Q06174|EST_BACST CARBOXYLESTERASE PRECURSOR (EC 3.1.1
25 .1). >pir:pir|JC1374|JC1374 carboxylesterase (EC 3.1.1.1) -
Bacillus stearothermophilus (strain IFO 12550) >gp:gp|D12681
|BACPBH7_1 Bacillus stearothermophilus esterase gene. NID: g
216313.
atgcaaatataaactacccaaaaccattctttttgaagaagggaacgtgcagtggtactt
30 ctacacggctttacaggttaactctgctgatgtaagacaacttgggcgttatcttcaaaaa
aagggtctatacatcttatgctccacaatatgaaggacatgcagcgccccagaagaaata
ttaaactctagccctttgtttggtttaaagatgttttagatggttatgattatttagta
gatcaaggttacgaagaaatagcagtagctggtttatcattaggtggcgcttcgcatta
aaactaagtttaaatcgtgatgtgaaggggattataactatgtgtgcacctatggagaat
35 aaaacagaaggttcgatttatgaaggctttcttgaatatgcacgtaactttaaaaaatat
gaaggcaagatcaacaaacgattgatcaagaaatggaacaatttcacccaactgaaacc
ctgagagaactgagtgacactctaaatggagttaaagaacatgtcgatgaagtaattgat
ccaataacttgtcgtacaagcagaacaagatacaatgattgatcctcaatcagcaaatat
atatataatcatgtcgattctgatgaaaaagaaatcaaatggtatcaacattcaggtcat
40 gtgattaccattgataaaggcattgtcaaatcctcctgtaatttttagcgtatctggaa
aatcatcagcagtaa

Sequence 72

MQIKLPKPPFFEEGKRAVLLHGFNGSADVRLGRYLQKKGYTSYAPQYEGHAAPPEEI
45 LKSSPFVWFKVDLDGYDYLDVQGYEEIYAVAGLSLGGAFALKLSLNRDVKGIITMCAPMEN
KTEGSIYEGFLEYARNFKKYEGKDQQTIDQEMEQLFPTETLRELSDTLNGVKEHVDEVID
PILVVQAEQDTMIDPQSANYIYNHVSDEKEIKWYQHSQGHVITIDKGHCQNPPVILAYLE
NHQQ*

Sequence 73

Contig_0446_pos_947_468,
putative peptide of unknown function
55 gtgtcttcttcaataagtatatttatacatatctgtaataattgacggttctgaacctagc
ttgcctgcgaatgtgattttatcacctttttgcgctgccataggtatggcaatagctatg
ataatcacaattacaattgtccctaaagaaatgagcaatttttatatgataaacgttcc
atgtatcttaaaataaaatcaaaaataatagctagaagtgacgctggaatagcacctatt
aaaatgagtgacactattgttacgatcaatgcctaataatattaaatctcctagaccacca

gcgccattataaagctgcgagtgtagcagtagccaatgattaataccatagctgtgcgtatt
cctgccatgataacaggcattgcaatagggagttcgacttttggtcaatcttctaagtggg
ttcattccaatgccttttagccgcttcaataagagagggatcgacctccttaataaccttaa

5

Sequence 74

VSSSISILYIFVIIDGSEPSLPANVILSPFCAAIGMAIAMIITITIVPKEMSFLYDKRS
MYLKIISKIARSAGIAPIKMSALLRSPNNIKSPRPPAPIKAASVAVPMINTIAVRI
PAMITGIAIGSSTLVNLLSGFIPMPLAASIREGSTSLIP*

10

Sequence 75

Contig_0447_pos_18108_18413,

putative peptide of unknown function

atgtcttctcctcctatccacattaatattttgatagattgtagagtttaatttatctgga
15 acacgtaaatacatcaaaaactacacctaataatgttctttatatttgaagtcgttctctgtc
acctctttatcgaagaatttcagtgaaaccattatctttatgtctatttcctacaagtgt
ttaatgagggtagacttacctgaaccatttttcccaattaaacccactacctcgccaggt
ttaacagagaatgtgatgtcagtaagctggaattctgaattcttatatgatttattgagt
tgctga

20

Sequence 76

MSSHPIHINILIDCRVNLSGTRKSSKTPKCSLYLKSFVSLSKNFSEPLSLCLFPTSV
LMRVDLPEPFFPIKPTTSPGLTENVMSSWNSEFLYDLLSC*

25

Sequence 77

Contig_0447_pos_22881_22129,

putative peptide of unknown function

gtgaggcaaatgtcacagtatccacttttgaatcaattaaatactttaaaagaggctcag
tggtcgatttaacacatactttcgacccaaatatccctcgtttttagcgaatttgaaaaa
30 ggtgaagttctcaacgctattcaatgttaaagatcatgggttttatgtacaacgttggagt
atcgtaactcaatatggaacacacattgatgtccaatccatttcggtgaaaatagaaga
tatttagaagaattagattttaaagaacttgttttaccattaattgtttttagattattct
aaagaagctgcacaaaattcagattttatcgatcacgtaaacatttagaagattgggaa
caacaacacggctgcattgaagcaggtacttttgcgcattacgtactgattgggtcaaaa
35 cggtggccagatatagaaaaatttgaaaaataagatgtagatggccatcaacatcttcca
ggttggggccttgatgcattaaaatttctcattgaagaacgtgggtgtaaatccataggt
cacgaaacatttgatactgatgcctcaattgatacagctaaaaatgggtgatattgttggc
gaaagatatatcttaggtcaagacacattccaagtcgaattacttaccatttagatcaa
ttacctaccagaggtgcaattatctatgcaatcagcccaaaaccaaagatgcaccaggc
40 tttccagttcgtgcattcgcaataaaaaccttaa

Sequence 78

VRQMSQYPLWNQLNLKEAQWVDLTHTFDPNIPRFSEFEKGEVSTLFNVKDHGFYVQRWS
IVTQYGTHIDAPIHFVENRRYLEELDLKELVLPVLDYSKEAAQNSDFIVSRKHLEDWE
45 QQHGRIEAGTFVALRTDWSKRWPDIEKFENKDVGHQHLPGWGLDALKFLIEERGVSIG
HETFDTDASIDTAKNGDIVGERYILGQDTFQVELLTNLDQLPTRGAI IYAI SPKPKDAPG
FPVRAFAIKP*

50

Sequence 79

Contig_0447_pos_19018_18668,

putative peptide of unknown function

gtgtattttatttagtccaaaacgctcacttcactataaagctaaaattgacgcaaat
55 atttcagatgatttagcagatacatatgaaaataaatcatacatcaaatcattgaaagta
agattttatttacacaatgcaattaattgtcgctttttattgcaattttaatacccgtcata
ggaaatgcatctgagaatcacatcgctctaataatgattcctttcattattacaatcatt
tcatccataatgattgggatattttatagaaaatttgatgctcgataccctaaattagga
gagaaacgttacactgaaaaagcatttaattattatggacgaaggagagtga

Sequence 80

VYLFIVQNASLHYKAKIDANISDDLADTYENKSYIKSLKVRFIYTMQLIVAFIALIPVI
GNASENHIALIMIPFIITIISIMIGIFYRKFDARYPKLGEKRYTEKAFNIMDEGE*

5

Sequence 81

Contig_0447_pos_18433_18116,

putative peptide of unknown function

atggaaaacttattagaagttcagcaactcaataaatcatataagaattcagaattccag
10 cttactgacatcacattctctgttaaacctggcgaggttagtgggtttaattgggaaaaat
gggtcaggttaagtctaccctcattaatacacttgtaggaaatagacataaagataatggg
tcactgaaattcttcgataaaagaggtgacagagaacgacttcaaatataaagaacattta
gggttagtttttgatgatttacgtgttcagataaattaactctacaatctatcaaaata
ttaatgtggataggatga

15

Sequence 82

MENLLEVQQLNKSYPKSEFQLTDITFSVKPGEVVGLIGNSGSKSTLINTLVGNRHKDNG
SLKFFDKEVTENDFKYKEHLGVVFDLLRVPDKLTLQSIKILMWIG*

20 Sequence 83

Contig_0447_pos_18109_17288,

putative peptide of unknown function

atgttcaagaacaacaaaagtattgaagatacttatgcaacaaaacctattattcagaat
atcgttgggtcaggcacaaatcaacaagtgtatggcgaaacaaacacccatgagatatagc
25 ttgaaagctatcatggctggttttctattatcaatagttacagtttttatgttagcaatt
aaaacacaattcgcttcaacgcataatgacgggttaatacaatttgatgggagctattgag
tttagtttaggtctcgtatttagttgtttaaccaattctgaattattaactagtaatttt
atgtatctgactgttggttggattataaagcaattagtgttaagtaaaatgatatggatt
tttattttctgttttataggtaatatcttaggtggattttattttttttcctcatgaaa
30 tatgcacatgttatgacgccagaaatgacagatagtttaacagcatttagtacataaaaaa
acagtagaatcgacttggttaaataattttgattaaaggtatattttgtaattttctttatt
aatatcggtatttttatttcaatgcagtttaagagggactagccaaagcattctttata
gcttggtgagtgattgtctttgtattttatgggttacgaacacggttggttttaacgctgga
ttatatgcaggtatgatgttctttaatatggatggattatcttggttggtgtgtctaaaa
35 aatattgtttttgcattccttggaactatatcggtggaggtatctttattggattagt
tatgcatatttgaacggtaaacgtgacagcctccaaccatag

Sequence 84

MFKNNKSIEDTYATKPIIQNIVGQAQIKQVMAKQTPMRYTLKAIMAGFLLSIVTVFMLAI
40 KTQFASTHNDGLINLMGAI AFSLGLVLVLTNSELLTSNFMVLTGVWYKAVISVKMIWI
FIFCFIGNILGGFILFLMKYAHVMTPEMTDSLALVHKKTVESTWLNILIKGIFCNFFI
NIGIFISMQFKEGLAKAFFIACGVIVFVFMGYEHVVFNAGLYAGMMFFNMDGLSWLGVLK
NIVFAFLGNYIGGGIFIGLVYAYLNGKRDSLQP*

45

Sequence 85

Contig_0447_pos_16854_16309,

is similar to (with p-value 2.0e-31)

50 >sp:sp|P54951|YXEL_BACSU HYPOTHETICAL 19.0 KD PROTEIN IN IDH
-DEOR INTERGENIC REGION. >gp:gp|Z99124|BSUB0021_55 Bacillus
subtilis complete genome (section 21 of 21): from 3999281 to
4214814. NID: g2636442. >gp:gp|D45912|D45912_15 Bacillus su
btilis genome sequence between the iol and hut operon, parti
55 al and complete cds. NID: gl408482.

gtgaaaaatattgtcgaatattaatatttagagtggcacatgaacaagatgctgaagaatta
catagcatcatgcaaattgcttttacacctttaagagaactaggtattgattggccatca
gttcacgctgatcttgaaatggtaagagataatttaagacaaaatactacatttgactt
gaaaatgaaaaagaaattattttcaacgattacgggttgctatgcatggagtagtgtaaaa

cccatttcaggttatccggttcggttggtggttgaacacgaccaacttatgatggacaa
 ggggtatgggagtcacatttttaaaatatgtagaggagacatttttacgcgatactttaaaa
 gctgctgcggttaaccttaggaacatcagcacggtttgcacccttggttattaaacatttac
 gaaaagcggggttatgaaatatacgctaaacatgaaaatgatgatgggtgatttaggagtc
 5 ataatgcgtaaaattttaataaccagaacaatttaatatgatgacattttgggcccgcgcga
 ttttag

Sequence 86

VKNMSNINIRVAHEQDAEELHSIMQIAFTPLRELIGIDWPSVHADLEMVKDNLNQNTTFVL
 10 ENEKEIISTITVCYAWSSVKPISGYPFVWWFATRPTYDGGQYGSQLLKYVEETFLRDTLK
 AAAVTLGTSARLHPWLLNIYEKRGYEIYAKHENDDGLGVIMRKILIPEQFNDDILGRPP
 F*

Sequence 87

15 Contig_0447_pos_13818_13129,
 putative peptide of unknown function
 atgagaaaaggaaatcagaatgaagcttttagaagaatttatcggaactttattaaaagat
 gagcaatattatttaggttagcatttttagaaagtgaacacaaaatccttgaaatcata
 atggagaagatgattaagcaaggaattacaaaatttcgtattgtacctttactcattttt
 20 agtgcaatgcattatcagtgatattccacaaatacttaaagagatgaaagctcgatat
 ccacaaattgatagtaaaatgagtgcgccctcttggtacacatccatatatgaaaacatta
 gtagaaaatagaattgctgatgaaaaagtcagtgaaaggtcaaccaaagcaactatagta
 attgcccattggaatggaagtggacggttttacgaaagcacatgatgaattaaaagcattt
 gttaaaacgcttgatagtcacatcctgtttatgcaagagctttatatgggacattagca
 25 tttaaaatgatttagataaaaatctcagagcaaatatgacgagtttagtcattgtccatta
 tttttatttgatggtagattgggtgaataaagttaaagcgtcttttaggtgaaatgacattg
 catagtcattacacattacgccatcgattaactttgatccaattttaagattaattatt
 agagaaagacttgaagcggttagatatttaa

Sequence 88

30 MRKGNQNEALEEFITLLKDEQYYYELAFLESETQNLEIIMEKMIKQGITKFRIVPLLI
 SAMHYISDIPQILKEMKARYPQIDSKMSAPLGTHPYMKTLENRIADEKVSEGSTKATIV
 IAHGNGSGRFTKAHDELKAFVKTLDSHHPVYARALYGTAFKNDLDKISEQYDELIVIVPL
 FLFDGRLVNVKVRLLGEMTLHSQHLITPSINFDPILRLIIRERLEALDI*

Sequence 89

Contig_0447_pos_12691_10286,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P42435|NASD_BACSU NITRITE REDUCTASE (NAD(P)H) (EC 1.6
 40 .6.4). >gp:gp|D30689|BACNARB_4 Bacillus subtilis DNA around
 narB region (nasB operon and nasA gene). NID: g710016. >gp:g
 p|Z99105|BSUB0002_159 Bacillus subtilis complete genome (sec
 tion 2 of 21): from 194651 to 415810. NID: g2632457. >gp:gp|
 D50453|D50453_33 Bacillus subtilis DNA for 25-36 degree regi
 45 on containing the amyE-srfA region, complete cds. NID: g1805
 369.

atggcaaaacaaaaacttgtaatgattggtaatgggtatggcaggtttaagaacgatagaa
 gagatttttagaacgttcacaatcacaatttgatattactattattgggaaagaaccttat
 ccgaactataacagaattatggttatccaatattttacagaagaaaatgaccgtcgaagat
 50 acaattatgaatccttatgattgggtatcaagagaataatattgaacttataataatgat
 ccagtggaagaaagttagataaagaaaacaaaatagttactacttctaaaggtattgaagta
 gagtatgacatttgatttttcgctacttggtatcaaaagcctttgtattacctataaccagg
 tcaaatcttccttagtgatcattggatggcgaacaattgatgatacaaaataaaatgattgaa
 attgcccacacgaaaaacgcgcaggttgatcattgggtggaggtcttctaggcttagagtgt
 55 gccagaggacttctagatcaaggaatggaagtgcaggttcttcatttagctgattggctc
 atggaaatgcaattggatcgtaaagccggagaaatgcttaaagcggatttagaaaagcaa
 ggtatgaagattgaacttcaagcaaatctaaagaaatcattgggtgataaagatgttgaa
 gctattaaattagctgacgggtcggtgattgaaacagatttagtagttatggctgttggt
 atcagaccttatactgaagttgctaagatagtggttagatgtcaatagaggtattgtt

gtaaatgattatatgcaaacaatctgattctcatatttatgcagtcggtgaatgtgccgaa
 catgatgggaaagtttatggattggtggcgccactttatgaacaaggcaaagtgttagca
 gattatattaactggtaaaagaaacaaaaggtataaaggatctactactttcacttcactt
 aaagtatctggttggtgatttatatagtcagggcaaattggtgaagatgaagatgtccat
 5 ggtgtggaaatttttaaatagtgctgcacaatatctacaaaaaagtgtatttaagtcaggg
 caagtcggttggtgctgcttctgtatggtgatactgatgatggatcacgattttataatatg
 atgaaaaaacatgaaacgcttgaagattatacacttgtttctttattgcataaagggtgat
 gaagatgcggggacatctattgctgatatgtctgatgatgaaacgattttgtggatgtaat
 ggtgttgataaaggaacaatcgatcaatgctattacaagtaaaaggtttaacgctctgtagat
 10 gaagtgaactaaagcaacaaaagcaggttaattcatgtggtgaagtgtaaagggtcaaatcggt
 gagttattacaatatatacattaggtgacgactttattgctgcaaaaccaacaggtatttgt
 ccattgactgattttaacaagagaccaaattgtaactcaaatcagggctaaaaatctcaaa
 tcatcaaaagaagtacgacacgcttcttgatttcaaagataaagatggtgtcctaaatgt
 cgacctgcaattaattattattttaaatatggtttatccttttgacatcgagacgaaaaa
 15 gattctcgcttcgctaatgaaagatatcatgcaaatatacaaaatgatggtactttctca
 gtgattcctcaaatgcgcggtggtgttacagatgctgaccaactcattcgattaggagaa
 gttgctaaaaagtataacgtaccacttgttaaagtaacaggttcgcaacggttaggttta
 tatggattgaagaagaagaattaccacaagtttggaagatttaggaatgcgttctgct
 tctgcttatggttaaaagacgcgcttctgttaaaagttgcgttggttaagagttttgtcgt
 20 tttggtacacaatacacaaactcgactaggaataagacttgaaaaaacatttgaatatatt
 gatacacctcataaatttaaaatgggagatcaggttgtccgagaagttgtgtagagtct
 ggtgttaaagattttggcgtcatatctgttgaaaatggctacaaatatttatcgagggt
 aatggtggtactgatgttactgtaggttaattgttaacgacagttgaaaccgaagatgaa
 gtgattcaattatgtggtgccctcatgcagattacagagaaacaggtgtttacgctgaa
 25 gaacagcaccatggttagaacgtatggcctttgaaaatgtcaagaatgtcttattaaat
 caagaaaagcaaaaagaactgtatttaagaattatggaagccaaaaaagctgttgagaat
 gaaccatgggaaactattgttgaaaataaagaagcacaaaaaatctttgaagttgagaag
 gtgtaa

30 Sequence 90

.MAKQKLVLMIGNMAGLRTIEEILERSQSQFDITIIIGKEPYPNRYNRMILSNILQKKMTVED
 TIMNPYDWYQENNIELINNDPVEKVDKENKIVTTSKGIIEVEYDICIIFATGSKAFVLPPIG
 SNLPSVIGWRTIDDTNKMIEIAQTKKRAVVIGGGLGLECARGLLDQGMETVTLHLADWL
 MEMQLDRKAGEMLKADLEKQGMKIELQANSKEIIGDKDVEAIKLADGRVIETDLVVMVAVG
 35 IRPYTEVAKDSGLDVNRGIVVNDYMQTSDSHIYAVGECAEHDKVYGLVAPLYEQGKVLA
 DYLTGKETKGYKGGSTFTSLKVSGLDLYSAGQIVEDEDVHGVIEIFNSVDNIYKKVYLSQG
 QVVGAVLYGDTDDGSRFYNMMKKHETLEDYTLVSLHLKGGDEAGTSIADMSDDETI CGCN
 GVDKGTIVNAITSKGLTSVDEVTKATKAGNSCGKCKGQIGELLQYTLGDDFIAAKPTGIC
 PCTDLTRDQIVTQIRAKNLKSSKEVRHVLDFDKDKGCPKCRPAINYLLNMVYPFEHRDEK
 40 DSRFANERYHANIQNDGTFSVIPQMRGGVTDADQLIRLGEVAKKYNVPLVKVTGSQRVGL
 YGLKKEELPQVWDLGMRASAYGKKTRSVKSCVGKEFCRFGTQYTRLGIRLEKTFEYI
 DTPHKFKMGVSGCPRSCVESGVKDFGVISVENGYQIFIGNGGTDVTVGKLLTTVETEDE
 VIQLCGALMQYYRETGVYAERTAPWLERMGFENVKNVLLNQEKEKELYLRIMEAKKAVEN
 EPWETIVENKEAQKIFEVEKV*

45

Sequence 91

Contig_0447_pos_10283_9969,

is similar to (with p-value 3.0e-22)

>sp:sp|P42436|NASE_BACSU ASSIMILATORY NITRITE REDUCTASE (NAD
 50 (P)H) SMALL SUBUNIT (EC 1.6.6.4). >gp:gp|D30689|BACNAB_5 Ba
 cillus subtilis DNA around narB region (nasB operon and nasA
 gene). NID: g710016. >gp:gp|Z99105|BSUB0002_158 Bacillus su
 btilis complete genome (section 2 of 21): from 194651 to 415
 810. NID: g2632457. >gp:gp|D50453|D50453_32 Bacillus subtili
 55 s DNA for 25-36 degree region containing the amyE-srfA regio
 n, complete cds. NID: g1805369.

atgaaagctaaagaaaagattaaagttacaacaatgaatgaaatgattcctcaaataggc
 aaaaaagtagttgtaaacgaaaaagaaataggtatttttctcacagataatggtgattta
 tatgccattggaaatatatgtccacataaagaaggaccgttctgaagggactgtaagt

ggtgattatgtttactgtccgttacacgatcaaaaaatagctttaaaaactggagaagta
caacaacctgatacaggatgtgtagagacatacgaagtagaagttattgatggagatatt
tacttatgtctataa

5 Sequence 92
MKAKEKIKVTTMNEIPQIGKKVVVNEKEIGIFLTONGDLYAIGNICPHKEGPLSEGTVS
GDYVYCPLHDQKIALKTGEVQQPDTGCVETYEVEVIDGDIYLCL*

Sequence 93
10 Contig_0447_pos_9876_9043,
is similar to (with p-value 3.0e-32)
>sp:sp|P29928|SUMT_BACME UROPORPHYRIN-III C-METHYLTRANSFERAS
E (EC 2.1.1.107) (UROGEN III METHYLASE) (SUMT) (UROPORPHYRIN
15 OGEN III METHYLASE) (UROM). >pir:pir|A42479|A42479 S-adenosy
l-L-methionine uroporphyrinogen III methyltransferase - Baci
llus megaterium >gp:gp|M62881|BACCOBA_1 Bacillus megaterium
S-adenosyl-L-methionine:uroporphyrinogen III methyltransferas
e (COBA) gene, complete cds. NID: gl42694.
gtgatttttatttgatcgtctcgtataatcctttcatcttacagtatgcttcttctcaaaca
20 aaagtgatcaatgtgggaaagaaaccttattgtaaacacattcaacaagaggagattaat
caaaaaattgttgaaagcagctaataatcaatgtgtggtgagactaaaggaggagat
cctgcatgttttggtagaattacagaagaagtacaaacattagaaaatcatcatattcat
tagcagattgtccctgggtgtgacatcagcaagtgtgtagcaactatgaatatggga
ttaacgatgcttctatcgacccgagtgtagcttctcaactggtcattttaaagattcg
25 gttaatcacgatacggatatttaggaacttgattaatggaggcacttttagctatttatatg
ggtgtgaaaagattagggtcaaatatttaaacaaattgaatcatatacgaatgaagactac
cccattgcaatagtgtttaatgcttctgctacaatgaaaagattgttataggtcattta
agtagcattgaagaacaattgggttctcaacaactagaaggatccaggcatatgcatt
ttaggtaataacttgatgacattaatcgtagcttattgaataataataagaatgacaag
30 ggaaatctatatttaatacaaggagataaagaacgtgcaattgcaaaggctgaaacttta
tatgatgaaggaatccaatgtctgattgattttgaccatagctaccacatttctcaacaa
aacgtgtataacgaaatgattaaacacaagagtattaaacaatatatgtataa

Sequence 94
35 VILFDRLVNPFIQYASSQTKVINVGKKPYCKHIQEEINQKIVEAANQYQCVVRLKGGD
PAIFGRITEEVQTLNHHIHYEIVPGVTSASAATMNMGLTMRSIAPSVTFSTGHFKDS
VNHDTDIRNLINGGTLAIYMGVKRLGQIIKQIESYTNEDYPPIAIVFNASCYNEKIVIGHL
STIEEQLVSQLEHGPICILGNILDDINRTLLNNKNKNDKGNLYLIKGDKERAIKAETL
YDEGIQCLIDFDHSYHISQQNVYNEMIKHKSIIKTIYV*

40 Sequence 95
Contig_0447_pos_8864_5181,
is similar to (with p-value 0.0e+00)
>gp:gp|AF029225|AF029225_1 Staphylococcus carnosus NarG, Nar
H, NarJ, and NarI genes, complete cds. NID: g3929521.
atgggaaaatttgattgaatttctttaaaccgacagaaaagttaaaggaaattgggtcg
gtattagagcataaaagtcgagaatgggaaaagatgtatagagaaagatggagccacgac
aaagttgtgagaacgcgcgcatgggtgttaactgtactggatcatgttcatggaaagtattt
gtcaaaaatggcgtaattacatgggaaaatcaacaaattgattatccaagttgtggacct
50 gatatgccagaatttgagccaagaggtgtccgagaggtgcatcatttcttggtatgag
tatagtcggttaagagttaaatatccttatattagaggtaaattattagatttatggacc
gaagcgcttgaagaacaaaaaggaaccgaattgcggcatgggcatccatcgtagaaaat
gaagaaaaagccaaacaatatataaagaagcaagaggtaaagggtggacacgtcagagcaaat
tggaagatgccacagatatcattgcagctcaaattttatacaccataaaaaaagatgga
55 ccggatcgtattgtcgtgatttactcctattcctgctatgtcgatgattagttatgcttca
ggagcaagattttattaatgttaggtggagaaatgttaagtttttacgatttggtatgct
gatttaccacctgcatctccacaaatttggggtagcaaacagacgtgccagaatccagt
gattggtacaaacgctcatacataatgatgtgggatcaaacgttccattaacacgtaca
cctgacgcacattttatgactgaagtttagatataaaggggcgaaagttatttcagtagca

cctgattatgctgagaatgttaagttcgccgacattggcttgaccacatccggggaca
 gatgcagcgggtgcacaagcaatgacacatgttattttacaggaatattatgaaaatcaa
 ccgaatgatatgtttatctaactatgctaagcaatattctgatatgccgtttgtcattatg
 'ttagatgaagatgagaatggctataaagcaggtagattcttgctgcttctgatttaggg
 5 atgtcaggtgaaaataatgaatggaagccagttattcaagacaaattgagccaacaatta
 ctgttccctaattggcacaatggggcaacgctgggaagaaagggaaaaaatggaatttgaaa
 cttgaaacagaggatggtacaccaattgatccaatgttatcaatggttgaaagtgactat
 catgttgaaacgattcaatttccatattttgatagcagtggtgatggtatctttgagaga
 cctattgcaacgagaactattcagttagctaacggagagaagaagttaaaaattgctacggtt
 10 tatgatttaatgacgagtcattatggtgttcaacgttttgaacacgaactagaagctaca
 tcttatgatgacgcatcttctaaatatactcccgttggaagaacaaattacaggtatc
 aaaaaagaattagtgacgaaagtggcaaaagaatttgcaaaaaatgctattgatactggt
 ggacgctcaattgattattatgggggctggtatcaaccattgggttaactccgatactatt
 tatcggttcaatttcttaacttagtactattgtgtggttgcaaggcgttaacgggtggtggt
 15 tgggcacactatgtaggacaagaaaaatgtcgaccaattgaaggatggaatactattgca
 ttcgctaaagattggcaaggtcctccacgtttacaaaaatggtacaagttggttctatttc
 gctacagatcaatggaagtatgaagaatcaaatgtagataaattaaaaatcaccattagct
 gaaaataattaagcatcaacatccagctgattacaatgtaacagctgctcgtatgggctgg
 ttgccttcataatccacagtttaataaaaaacagctctattttggtgaagaggctaaagat
 20 gaaggtgatgattcaaatgaagccattctacaaaaagcgattgaatcagttaaaaataaaa
 gatacacaattttgcgatagaagatccagatttaagaaaaaacctcctaaaacattattt
 gtatggagatctaatttaatttctagttcagctaaaggacaagaatactttatgaagcac
 ttgttaggtgcgcgctctggttttaattggcagagccaaatgaagatgataaaccagaggaa
 attaaatggcgcgaggatacagaagggaaacttgatttattagtatcacttgatttcaga
 25 atgactgcgacgccatttatattcagatatcgttttacctgctgcaacttggtatgaaaaa
 catgatttatcttctacagacatgcatccatttattcatccatttaaccagcgattgac
 ccattatgggaatcgcgcttcggactgggatatttataaaactctaagttaaagctgtttca
 gaaatggcgaaagattatcttccaggtaaatttaaagatgtcgtaactacaccattagga
 catgatlcacaaacaagaaatttcaactgaatcgggtattgtaaaagattggtctaaagga
 30 gaaattgaaggtgtgccaggtaaaacaatgcctaatttttctatcgtagagcgagactat
 acacaaatttacgataaattcggttactgttggtccaaaactagaaaaagggaaaaataggt
 gctcatggtgtgagttatagcgttagtgaaagtagcgaagaacttaaaagtatagttgga
 acttggatgatgataatactattttagttaaataatgatagaccgagaatagatacagcg
 agaaaagtagcagatgtcattttgaatatatcctctgctacaaacggcaaatatcacaa
 35 aagtcatatgaagatttagaaaaatcaaacaggtatggaacttaagatatttctaaagaa
 cgtgcttctgaaaagatatcatttcttaaacattacttctcaaccaagagaagtgattcca
 actgcagtatccctggctctaataaagatggaagacgctactcaccgtttacaactaat
 gttgaacggttagtgccatttagaacactaactggacgtcaaagttattatatagatcat
 gaggtattccaacagtttggcgaaagtttaccggtatataaacctactttacctccaatg
 40 gtatttgggtgctcgtagataaaaaagtttaagggtggacaagatacattagtgttcgatac
 ctacacactcatggaataatggaatattcattcaacttatcaagataatgaacgcatgttg
 acgttggttagaggtggaccagttgtatggatttcaaatgaagacgcagctgacctggt
 attaatgataacgactggttagaagtatacaacagaaacggagttgttactgccagagct
 gtaacatctcatcgtagcctagaggcacaatgtttatgtatcatgcacaagataaacat
 45 atagagacacctggttctgaaattactgatactcgtggaggttctcataatgcacctact
 cgtattcacttgaaacctactcaattagtaggaggatagcacaattagttatcacttt
 aactattatggaccaattggaatcaaagagatgagtagtagctgttagaaaaatgaag
 gaggtcaattggccttgaagattaa

50 Sequence 96

MGKFGNLNFFKPTKEFNGNWSVLEHKSREWEKMYRERWSHDKVVRTTHGVNCTGSCSWKVF
 VKNGVITWENQIDYPSGCPDMPFEFPRGCPRGASFWSWEYSPLRVKYPYIRGKLDDLWT
 EALEEQKGNRIAAWASIVENEEKAKQYKEARGKGGHVRANWKDATDIIAAQILYTIKKG
 PDRIAGFTPIPAMSMISYASGARFINLLGGEMLSFYDWYADLPPASPQIWGEQTDVPES
 55 DWYNASYIMMWGSNVPLTRTPDAHFMTEVRYKGAKVISVAPDYAENVKFADHWLAPHPGT
 DAAVAQAMTHVILQEYYENQPNDFINYAKQYSDMPFVIMLDEDENGYKAGRFLRASDLG
 MSGENNEWKPVIQDKLSQQLLVNGTMRQWEEGKKNLKLETEDGTPIDPMLSMVESDY
 HVEITQFPYFDSSGDI FERPIATRTIQLANGEVVKIATVYDLMTS QYGVQRFHEHELEAT
 SYDDASSKYTPAWQEITGIKKELVTKVAKEFAQNAIDTGGRSMIIMGAGINHWFNSDTI

YRSILNLVLLCGCQGVNGGGWAHYVGQEKCRPIEGWNTIAFAKDWQGPRLQNGTSWFYF
 ATDQWKYEESNVDKLSPLAENIKHQHPADYNVTAARMGWLPSYPQFNKNSLLEGEAAKD
 EGDDSNAILQKAIESVKNKDTQFAIEDPDLRKNHPKTLFVWRSNLISSSAKGQEYFMKH
 LLGARSGLMAEPNEDDKPEEIKWREDTEGKDLLVSLDFRMTATPLYSDIVLPAATWYEK
 5 HDLSSTDHMFPIHPFNPAIDPLWESRSDWDIYKTLKAVSEMAKDYLPGKFKDVVTTPLG
 HDSKQEIISTEYGIVKDWKGEIEGVPGKTMNFSIVERDYTIYDKFVTVGPKLEKKGIG
 AHGVSYSVSEEEYEELKSIVGTWNDNTISVKNDRPRIDTARKVADVILNISSATNGKLSQ
 KSYEDLENQTMELKDISKERASEKISFLNITSQPREVIPTAVFPGSNKDGRRYSPTTN
 VERLVPFRTLGTGRQSYIIDHEVFQQFGESLPVYKPTLPPMVFGARDKKVKGQDTLVLR
 10 LTPHGKWNHSTYQDNERMLTLFRGGPVVWISNEDAADHGINNDWLEVYNRNGVVTARA
 VTSHRMPRGTMFMYHAQDKHIETPGSEITDTRGGSHNAPTRIHLKPTQLVGGYAQISYHF
 NYYGPIGNQRDEYVAVRKMKEVNWLED*

Sequence 97

15 Contig_0447_pos_5167_3638,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF029225|AF029225_2 Staphylococcus carnosus NarG, Nar
 H, NarJ, and NarI genes, complete cds. NID: g3929521.
 atggtattgaatctagacaaatgtattggtgtcacttgcagtgatgacatgtaaaac
 20 acatggacaaatcgacctggtgcagaatataatgtggtttaataacgtagaaacaaacccg
 ggtgtaggatatccaaaaagatgggaagaccaaggacaatataaaggtggttgggtgcta
 aataaaaaaggaaagcttgaattaaaatctggtaacagatggtcaaaaattgctttagggt
 aaaatcttctataatccagacatgccactcattcaagattattatgaaccgtggacatat
 aactatgaacacttaaccaatgctaacaaggacagcactctcccgtaggcagactcac
 25 tctttaatttcagggtgatagattgaatcttaaatgggggccaactgggaagatgattta
 gctggagggtcacattacaggaccagaggatccaaatattcagaaaatagaagaagattat
 aaattccaatttcgatgagacatttatgatgtatttaccagactatgtgaacactgttta
 aatccaagttgcgtagcatcttgtccatcaggagctatgtataaacgagatgaggatggt
 atcgtactcgtcgatcaagaagcctgtcgaggttgagatactgtatgactggatgtccg
 30 tataaaaaagtatattttaactggaaaacgaataaagctgaaaaatgtacattttgtttc
 ccacgaatcgaagctggtatgccaaactgtttgttccgaaactgtacaggacgtatgaga
 tatttaggtgttttatatatgacgcagatcgcggttcaagaagcggcttcagctaaagat
 gaaaaagacttatacgaaaaacaattagacctattccttgatccatttgatgaggaagtc
 attgcacaagctgaaaaagatggaataaatcaagaatggattacagcagctcaaaactca
 35 ccagtgatataaattggcaatagaatataaaatggcctttccattgcacactgaatttaga
 actatgccgatggtgtggtattgtccacctttaagtcctattatgagttatttcgaaggt
 gaaaaatgcaggtcaaaatccagatatgattttccagctattgaggaaatgcgtttacct
 attcaataacttagcaaatatttaactgctggcgacacaaaacctgttaaaagagggtta
 caaaaaatggcgatgatgagaagttatatgcgttctcaataaacaacaaacctttcgat
 40 acttctaattagaacgattaggacttactgaaagacagatgactgaaatgtatcgctta
 ctagggtattgctaataatgaagatcggtttgtgtgccttcttccataaaagaacatat
 ttagatacttataaagcgcaaggaagtcaaggttacgggtggagagtactttggctcta
 tgtgaaggttgtggtgtgcagttcaatcaggtaaaactggacaagaatttataatgaa
 aatttctatggaggatcttccgtgattaa

Sequence 98

MVLNLDKICIGHTCSVTCKNTWTNRPGAEMYWFNNVETKPGVGYPKRWEDQGYKGGWVL
 NKKGKLELKSGNRWSKIALGKIFYNPDMPLIQDYEPWTYNYEHLTNAKQGQHSVATAH
 SLISGDRNLNKGPNWEDDLAGGHITGPEDPNIQKIEEDIKFQFDETFMMYLPRLCEHCL
 50 NPSCVASCPSGAMYKRDEGDIVLDQEACRGWRYCMTGCPYKKVYFNWKNKAKEKCTFCF
 PRIEAGMPTVCSETCTGRMRYLGVLLYDADRVQEAASAKDEKDLYEKQLDLDFDPFDEEV
 IAQAEKDGINQEWITAAQNSPVYKLAIEYKMAFPLHPEFRTMPMVWYCPLSPIMSIFEG
 ENAGQNPDIMFPAIEEMRLPIQYLANLLTAGDTKPVKEGLQKMAMMRSYMSQITNQPF
 TSKLERLGLTERQMTMYRLLGIKYEEDRFVVPSSHKETYLDTYKAQGSQYGGGEYFGSN
 55 CEGCGVAVQSGKTGQEIYNENFYGGIFRD*

Sequence 99

Contig_0447_pos_3585_3070,
 is similar to (with p-value 5.0e-62)

>gp:gp|AF029225|AF029225_3 Staphylococcus carnosus NarG, NarH, NarJ, and NarI genes, complete cds. NID: g3929521.
 atgaattttccagaaaaaatgacatttcacaaaaatatttgaagaaactatttctaag
 tctcaccgccggtacgaagatttgcttgcatatagagaagtcgatgaattatactttg
 5 tcagaaattaaagctatctatacagatacatttgatttttagtaaaaaacaccactctat
 atgacatttaataaatttgacacgcaaaaaggaacggggtcaaatgctagctaaattaaag
 gttttatagcaaatgtttggactaaaaatggttgataatgaattatctgattttctccca
 ttgatgctacagtttttgaagttgctgatttttaaaaatgatagtcgagcacaggaaaac
 cttcaacttgtcattatgattattgaagatggtacgtatgaaatggcaaataccctagct
 10 gaaaacaataatccctatgcatatgttgtcagtgcatgaagaaaaacgtaaaagcgtgt
 atcgtgcctttgaaagaggtggaatcatgcttaa

Sequence 100

MNFPEKMTFHPKIFEEETISKSHPGYEDLLAYREVMNYTLSEIKAIYTDTFDFSKKHPLY
 15 MTFNKFDTQKERGQMLAKLVLYEMFGLKMVDNELSDFLPLMLQFLQVADFKNDSRAQEN
 LQLVIMIIEDGTYEMANTLAENNNPYAYVVSALRKTLKACIVPLKEVENHA*

Sequence 101

Contig_0447_pos_2768_2400,
 20 is similar to (with p-value 2.0e-50)
 >gp:gp|AF029225|AF029225_4 Staphylococcus carnosus NarG, NarH, NarJ, and NarI genes, complete cds. NID: g3929521.
 atgttggtgctaactttaagacgactatccatcaaaaacgtagacgattaagttcattt
 tcagatatatttgtgaatatcgttttgttgattattttaaatgggttgattctacg
 25 cttgtaaccaatgcgattcaacctgaatttgattatcgtaaacattgcgatatgggtt
 agacatttattcatgttttctccaaatgctgacttaatgttaaacgtgccttggtcggtt
 aaactgcacatattattagggtttacagtgtttgctggttgccatttactcgtttagta
 catggttgaggtgtaccactgtcttatatgaacagaagatatattgtttatcgcaaaaac
 30 aaaattttaa

Sequence 102

MLLLTLRRLSIKNVRLSSFSDFVNIIVLLIILIMGCVSTLVNIAIQPEFDYRQTIAIWF
 35 RHLFMFSPNADMLNVPWSEKFLHILLGFTVFACWPFTRLVHVWSVPLSYMNRRIYVYRKN
 KI*

Sequence 103

Contig_0447_pos_2341_1928,
 40 putative peptide of unknown function
 atgaatttagataagttgagagcacaagagggttatgattttggtggtatcgctttatat
 gattatcatcacacttcacaccaattaaatggcaatatgtttcaggtaacacaaatgat
 agatataaaacttatcattttgagaaagggtagagggcttgctggaatggtgatgaaaacc
 ggtaagcgtatgggtattgctgatgtagatacagctttatctccagaagagaaagttaa
 45 tttccaatcattcttagtgagtcattgacagctgtagttgcagtcctttatgggttagaa
 aattcaatgtatggcgttttattattaggtcaaagaaatcatcagccgttacctcagtc
 ttggaccaacttaattgaaaaacaaatcggtatttttacagaataaaactag

Sequence 104

MNLDKLRQAEGYDFGGIALYDYHHTSSPIKWQYVSGNTNDRYKLIILRKGRGLAGMVMKT
 50 GKRMVIADVDTALSPEEKVKFPIILSESLTAVVAVPLWLENSMYGVLLLGQRNHQPLPQS
 LDQLNIEKQIGIFTEIN*

Sequence 105

Contig_0447_pos_1927_884,
 55 is similar to (with p-value 1.0e-29)
 >sp:sp|P54663|DEGS_BACBR SENSOR PROTEIN DEGS (EC 2.7.3.-). >
 gp:gp|L15444|BACDEGSU_1 Bacillus brevis protein kinase (degS
) gene, complete cds; transcriptional activator protein (deg

U) gene, complete cds. NID: g710494.

gtggtaaataatgttggagcaaaactgattttaagtttagagcaattactttaagaattattat
gaaaccacgaacgagaaaaattgtatttgttaatagacaaggcaaaattattgctatgaat
gacgcagcaaaagataattttaactgaggaagataattataatgctatgacaaatgcgatt
5 tgtcatcgatgcgaaggatactctaataatgatgtacaatcgtgtaagattgtttt
ttagagacaacgcaattacaacattccaatttccaagtatttatgaagacaaaagataat
gaaattaaagccttttacagctatgtatcaaaatattgatgaacaaagaggtattagtga
tttaccttacagaatgtggcgctcagattgaaaggcaagaaaaatgtatcaacaaaaa
atgttacatcgttcaattcaagcacaagaaaatgaacgaaagcgattttctagagaatta
10 catgatagtgtataacagcatatgctcaatatagatgttgaactaaggccttttgaagtat
aagcacaggataaggtgttagctgaaacatctcaacgtatagaaggccttattatcacag
cttattgatgatattagaaatatgtctgtgaattaagaccttcttctctcgacgattta
ggcattgaagcgtttttaaatacatattttaaacagtttgaagaaaattatggtatgcat
attaaatatgattcgaacattaaaggcatgcgttttgataatgaaattgaaacagttgtg
15 tatcgtgtgatttcaagagggtgtattttaatgctctaaaatatgctgaggttaataaatt
gaggttaagtacgcatagtgatggcaagcagctttagcagagggttgggatcgaggtaaa
gggttttagtttagatcatcacctaaaggctctggacttgattgtacggaatgagagaa
cgtgcagaattagttaacggtcatgttaatatagagacacataattaatagaggtaata
attacattagatataccgatttaa

20 Sequence 106
VUNMLEQTDLSLEQLLNYYETTNEKIVFVNROGKIIAMNDAKDILTEEDNYNAMTNAI
CHRCEGYSNEYDVQSKDCFLETTQLQHSNFQVFMKTKDNEIKPFTAMYQNIDEQRGISA
FTLQNVAPQIERQEKMYYQKMLHRSIQAQENERKRISRELHDSVIQDMLNIDVELRLKY
25 KHRDKVLAETSQRIEGLLSQLIDDIRNMSVELRPSSLDLGLIEAAFKSYFKQFEENYGMH
IKYDSNIKGMRFDNEIETVVYRVVQEGVFNALKYAEVNEIEVSTHSDGKQLVAEVVDRGK
GFSLDHHPKSGSLGLYGMRRERAEVLNIGHVNIETHINRGTIITLDIPI*

30 Sequence 107
Contig_0447_pos_855_199,
is similar to (with p-value 5.0e-42)
>gp:gp|AL034446|SC1A9_26 Streptomyces coelicolor cosmid 1A9.
35 NID: g4007685.
gtgaaaatagttatagcggatgaccatgcagttgttaggacaggattttcaatgatatta
aattatcaagaagatatggaagttgttgcaactgcagctgacggggtgaaagcttatcaa
aaagtgttagaacatcgaccagatgttttaatttttagatttgagcatgccgccaggagag
tcaggcttaatcgcaaccagtaaaatttctgaaagtttctctgatactaaaattttaata
40 cttacgatgtttgatgacgaagaatatttatttcattgtgttaaaaagtggtgctaaggga
tacatttttaaaaaattcacctgatgagcaattaatattggccgtacgtacagtatatcaa
gggtgaaacttatgttgatatgaaattgacgacgtcttttagtcaatgagtttgtcaatcaa
tcacaaacggatgaagtgtcatcatcttcagatccatttaaaattttatcgaaacgagag
ttagaaatattacctcttatagcaaaaggctatggcaataaagatattgcagaaaagttg
45 tttgtatcggtgaaaacggttagaggcacataaaacgcataattatgacgaaactaaattta
aagagtaaactgaattagttgaatatgccttaagaaaaaattattagaatttttaa

Sequence 108
VKIIVADHVVRTGFSMILNYQEDMEVVATAADGVEAYQKVLEHRPDVLILDLSPMPGE
50 SGLIATSKISESFPDTKILILTMFDDEEYLFHVLKSGAKGYILKNPDEQLILAVRTVYQ
GETYVDMKLTTSLVNEFVNQSQTDEVSSSDPFFKILSKRELEILPLIAKGYGNKDIAEKL
FVSVKTVEAHKTHIMTKLNLKSKPELVEYALKKKLLEF*

Sequence 109
55 Contig_0448_pos_2830_4107,
is similar to (with p-value 2.0e-74)
>sp:sp|P13702|MVA_PSEMV_3-HYDROXY-3-METHYLGLUTARYL-COENZYME
A REDUCTASE (EC 1.1.1.88) (HMG-COA REDUCTASE). >pir:pir|A44
756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88)

- *Pseudomonas* sp. >gp:gp|M24015|PSEHMGCOA_1 *P. mevalonii* HMG-CoA reductase (*mvaA*) gene, complete cds. NID: g151258.

atgaaaagtttagataaaggatttagacatttaacacgaaaagataaattaaaaaaactt
 gttgaatacggttggctagatgatgaaaactatgaaatattacttaatcatccgttaatt
 5 aatgaggaagtcgcaaacagtttaattgaaaatgtcattggtcaagggtgcaactaccagta
 gggttattacctcgaattatagttgatgataaagaatatgtagtacctatgatggtagag
 gaaccttctgtcgtagcagcagcaagttatggcgcaaaactcgttaatcaaagtggtagga
 ttaagacaatttcaagtgaacgtctaattgattggacaaattgtctttgatgatgttgaa
 gacacaggcacatttagctaactcaatatatacaaatagaatcacaattcatcaaactcgt
 10 gatgaagccttacccttctattaaagcaagaggtggaggatatcaacgtattgaaatagat
 acattccctaatacatcgattattatctttgaagggtttttgttgatactaaagatgctatg
 ggtgctaataatgttaaatacaataattagaagcaatcactgcacatctaaaagttaaattt
 tcaaatcaaaatgttttaatgagattttatctaatacatgcgacagcatcagtagtaaaa
 gtacaaggggaaatagatatattgaagatttacatagaggagagagaagtggcgaagaggta
 15 gcacaacgtatggaacgagcgtcagttcttgcaacagtagatatacatcgtgctgcaaca
 cataacaaaggtgtgatgaatggtatacacgctgtagtattgggtacaggcaatgatata
 agaggagttgaagcaagtgtcatgcatatgcaagcaaatggtcattatagaggagata
 gctacttgggaatatgatcgctcacgtaataaattggttggaaactattgaagttcctatg
 actttagcgacagttaggtggaggtacgaaagttttacctattgctaaagcctcattaaat
 20 ttgcttaatgttgaataatgcacaggaactagggaagttgttgctgctgttggttagca
 caaaatttctctgcatgtagagcgttagtgctgaggggatacaacaaggacatatgagt
 ttacaatataaatcattagcgattgttgtaggtgcaaaaggcgaagaaattgcgcaagta
 gctgaagcgtcacaatatgaatcacaagctaataactgccaaagctcaagaaatcttgatg
 aatataagaaagtcataa

25

Sequence 110

MKSLDKGFRHLTRKDKLKKLVEYGLDDENYEILLNHPLINEEVANS LIENVIGQALPV
 30 GLLPRIIVDDKEYVVPMMVEEPSVVAASYGAKLVNQSGGFKTISSERLMIGQIVFDDVE
 DTGTLANSIYQIESQIHQIADEAYPSIKARGGGYQRIEIDTFPNHRLLSLKVFVDTKDAM
 GANMLNTILEAITAHLKVKFSNQNVLSILSNHATASVVKVQGEIDIEDLHRGERSGEEV
 AQRMERASVLAQVDIHRAATHNKGVMNGIHAVVLATGNDTRGVEASAHAYASKDGHYRGI
 ATWEYDRSRNKLVTIEVPMTLATVGGGTVKLPIAKASLNLNVENAQELGQVVAAVGLA
 35 QNFSACRALVSEGIQQGHMSLQYKSLAIVVGAKGEEIAQVAEALKYESQANTAKAQEILM
 NIRKS*

Sequence 111

Contig_0448_pos_4618_4187,
 40 is similar to (with p-value 2.0e-20)
 >gp:gp|U96107|SCU96107_3 *Staphylococcus carnosus* N5, N10-meth
 ylenetetrahydromethanopterin reductase homolog, *SceB* precurs
 or (*sceB*) and putative transmembrane protein genes, complete
 cds, and putative Na⁺/H⁺ antiporter *NhaC* (*nhaC*) gene, parti
 45 al cds. NID: g2735503.

atgaaattcaaaaaattattatctcgtattattatcgctacaatgattacatttactgga
 acactctcatatcaagctattgaacaaacgcataatttcccatgctgcacataattattat
 ggtaaaaaacaatgcacttgggtgggcatttaaacgtcgtgctcaattaggtaaacctgta
 tcaaatcgatggggtaattgctaagaattggtatagcaatgcacgtcgatctggttatgca
 50 actggacataagcctcgaaaatacgcgtgttatgcaatcaacgagaggctattatgggcac
 gtagcagtggttgaaaaaagtatataagaatggaaaaatcaaaatttctgaatataattat
 aatgtgccattaggctacggcacacgcattattagtaaatcgtctgcacgaaactataat
 tatattttattaa

Sequence 112

MKFKKLLSRIIIATMITFTGTL SYQAIEQTHISHAAHNYGKKQCTWWAFKRRQLGKPV
 SNRWGNAKNWYSNARRSGYATGHKPRKYAVMQSTRGYGHVAVVEKVYKNGKIKISEYNY
 NVPLGYGTRIISKSSARNYNYIY*

Sequence 113
 Contig_0448_pos_2534_1422,
 is similar to (with p-value 2.0e-31)
 >sp:sp|P40830|PKSG_BACSU PUTATIVE POLYKETIDE BIOSYNTHESIS PR
 OTEIN PKSG. >gp:gp|U11039|BSU11039_2 *Bacillus subtilis* W168
 polyketide synthase (pksX and pksorfx6) genes, complete cds.
 NID: g602656. >gp:gp|Z99112|BSUB0009_183 *Bacillus subtilis*
 complete genome (section 9 of 21): from 1598421 to 1807200.
 NID: g2633902. >gp:gp|Z99113|BSUB0010_7 *Bacillus subtilis* co
 mplete genome (section 10 of 21): from 1781201 to 2014980. N
 ID: g2634090.
 atggcctaaacttgacagaagcgcgcgaagtcgacctaataaatttttaattggaattggt
 caaactgaaatgactgtgagcccagtgaaatcaagatatcgatatctatgggagccaatgct
 gctaaagatattataacagaagaagataaaaaagaatattggtatggttatagtagcaact
 15 gagtctgcgattgataatgccaaagcagcagccgttcaaattcaccatcttttaggtatt
 caaccctttgcaagatgctttgaaatgaaagaggcttgttatgcagcaacacctgcaatt
 caacttgccaaagattatcttgctcaacgcctaacgaaaagggttcttgctcattgctagt
 gacacagctcgttatggtattcattctggtggtgagcctactcaagtgccggtgcagtt
 gcaatgatgatttccataaaccgaagtattttaaaacttaatgatgatgcccgtagcatat
 20 actgaagacgtttatgatttctggtcccaacgggtcatcaatatcccttagttgctggt
 gcattgtcgaaagatgcctatatcaagtcattccaagaaagttggaatgaatatgcacgt
 cgccataataaaacactcgtgatttgcgttcactatgtttccatgtaccattcaccaaa
 atgggacaaaaagctttagattctattattaatcatgccgatgaaactacacaagaccgt
 cttaaactctagttaccaagatgcagttgattataatcgttatgtcggtaatatttacaca
 25 gggtccttatatttaagtctcattctctttattagaacacgtgattttaaaggcgacaa
 acgattggtctctcttagttatggttctggttctgtaggcgagttcttttagtggaacatta
 gtagatggattcaaggagcaatttagatggtgagcgccacaaatttttataaataataga
 atagaggtttctgttgatgaatatgaacatttcttcaaacgctttgaccaattagaattg
 aatcatgaacttgaaaaatcaaatgcagatcgtagacattttctattttaaactctattgat
 30 aacaatattcgtgaatatcatatagcagaataa

Sequence 114
 MAKLAEARQVDPNKFILIGIQTEMTVSPVNQDIVSMGANAAKDIITEEDKKNIGMVIVAT
 ESAIDNAKAAVQIHLLGIQPFARCFEMKEACYAATPAIQLAKDYLAQRPNEKVLVIAS
 35 DTARYGIHSGGEPTQGAGAVAMMISHNPSILKLNDNAVYTEDVYDFWRPTGHQYPLVAG
 ALSKDAYIKSFQESWNEYARRHNKTLADFASLCFHVPTKMGQKALDSIINHADETTQDR
 LNSSYQDAVDYNNRYVGNIYTGSLYLSLISLLETRDLKGGQTIGLFSYSGSGSVGEFFSGTL
 VDGFKEQLDVERHFKLLNNRIEVSVEYEHFFKRFQLELNHELEKSNADRDIFYLKSID
 NNIREYHIAE*
 40

Sequence 115
 Contig_0449_pos_584_919,
 is similar to (with p-value 3.0e-38)
 >sp:sp|P42874|URE2_STAXY UREASE BETA SUBUNIT (EC 3.5.1.5) (U
 REA AMIDOHYDROLASE). >pir:pir|S38484|S38484 urease (EC 3.5.1
 45 .5) beta chain - *Staphylococcus xylosus* >gp:gp|X74600|SXUREA
 BC_2 *S.xylosus* gene for ureA, ureB, and ureC genes for ureas
 e gamma, beta and alpha subunits. NID: g410513.
 gtgattgaagtaaaaaatacaggcgatagacctatacaagtaggttcacatttccacttt
 50 ttggaagcaataaagcattagaatttgatcgtgagaaagcatatggtaaacatttgat
 attcctgcaggagctgcagtgagatttgaaacctggagatgaaaaaaagtacaacttgct
 gaattattctggacgacgtaaaatttatggattccgtggttttagtcgatggcgatattgac
 gaagaacgcgtattccgtccaaatgattcaaatcaaaacgcgcgcttaaaaacgatgca
 55 ggcgaagacaatgcgaataaaaaagggtggttaataa

Sequence 116
 VIEVKNTGDRPIQVGSHFHFEEANKALEFDREKAYGKHLDI PAGA AVRFE PGDEKKVQLV
 EYSGRRKIYGFRLVDGDI DEERVFRPNDSNQNA AVKNDAGEDNANKKGK*

Sequence 117

Contig_0449_pos_922_2637,

is similar to (with p-value 0.0e+00)

5 >sp:sp|P42873|URE1_STAXY UREASE ALPHA SUBUNIT (EC 3.5.1.5) (UREA AMIDOHYDROLASE). >pir:pir|S38485|S38485 urease (EC 3.5.1.5) 62K chain - Staphylococcus xylosus >gp:gp|X74600|SXUREA BC_3 S.xylosus gene for ureA, ureB, and ureC genes for urease gamma, beta and alpha subunits. NID: g410513.

10 atgagtttttaaatgacacaatctcaatacacaagtctttatggaccaactgtaggagac
tctgtgagattagagatacgaacttgtttgcacaagttgaaaaagactatgcaaattat
ggagatgaagctactttcgggtggcggaataatcaattcgtgatggtatggctcaaaatcct
aatgtgacaagagatgataaaaatgtagccgatttagtttaactaacgcattaattatt
gattatgacaagattgttaaagcagatatcggaattaaaaatgggttatatttttaagatc
15 ggtaaagctggaaaccagatataatggataacgttgacatcatcattggtgcaacaact
gatattattgctgctgaaggtaaaattgttactgccggcggtatcgatacacacgtgcac
ttcatcaatcctgaacaagctgaagttgacttgagagtggtattacaacgcataatcggt
ggaggaactgggtgcttctgaaggtgctaaagcgactactgtaacaccagacacttggtcat
attcatcgcatgttagaagcagcagaagagatgcctattaatgtaggatttactggtaaa
ggtaagctgtcaatcatactgcacttattgaacaaattcatgcaggcgctatagggtcct
20 aaagtacatgaagattggggagctacaccttcagcattaaagtcattgacggttgca
gatgagtttgatgttcaagctcgctttacatgcagatacattaaatgaagctggatttatg
gaagatacaatggctgctgtgaaagatcggtgattgcatatgtatcatactgaaggagct
ggtggtggtcatgcacctgacttaatacaatcagctgcatattcaaacatcttaccttct
tctacaaacccaacattaccttacactcacaacactgtagatgaacatttagacatggtt
25 atgattactcaccatcttaattgcttcaataccagaagacattgcatttgcagatttctcgt
atacgtaaggaaactatagcagcagaagacgtattacaagatatggcggtatttagtatg
gtaaagttcagattcacaaagcaatgggacgtgctcggtgaagttgtaaacacgtacttgcaa
gttgacacacgtatgaaagaacaacgcggaccattagatgggtgactttgaatatcacgat
aataatcggtattaaacgttacattgcaaaatatacaatcaatcctgccattacacatggt
30 atttctgactatggttgatctgtagaagcgggtaaaacttgccgatttagtaatgtgggaa
ccagaattcttcgggtgcaaaacccgatcttgttggttaaaggtggcatgattaactcagca
gtaaatgggtgatgctaattggctccataccaacatcagagcctttgaaatatcgaaaatg
tatggtcaatttgggtgtaacattacacatactgctatgacttttgttttaacactgca
tatgaaaacggtattttatcgctcaactcaatctaaaaacgaatggttcgaccagttagaat
35 attagaaatttaactaaggcagatatgaaaaataataatgctacacctaataatagatgta
gatccacaaacatatgaggtatttcgttgatggtaataaaatcacaagtgaagcagcaaca
gaattaccattaacacaaagatacttcttattctag

Sequence 118

40 MSFKMTQSQYTSLYGPTVGDVRLGDTNLFQAQVEKDYANYGDEATFGGKSI RDGMAQNP
NVTRDDKNVADLVLTNALIIDYDKIVKADIGIKNGYIFKIGKAGNP DIMDNVDIIIGATT
DIIAAEGKIVTAGGIDTHVHFINPEQAEVALESGITHIGGGTGASEGAKATTVTPGPWH
IHRMLEAAEEMPINVGFTGKGQAVNHTALIEQIHAGAIGLVHEDWGATPSALSHALDVA
45 DEFDVQVALHADTLNEAGFMEDTMAAVKORVLHMYHTEGAGGGHAPDLIKSAAYSNI LPS
STNPTLPYTHNTVDEHLDMMVITHHLNASIPEDIAFADSRIRKETIAAEDVLQDMGVFSM
VSSDSQAMGRVGEVTRTWQVAHRMKEQRGPLDGFYHDNNRIKRYIAKYTINPAITHG
ISDYVGSVEAGKLADLVWWEPEFFGAKPDLVVKGGMINS AVNGDANGSIPTSEPLKYRKM
YGQFGGNITHAMTFVSNTAYENGIYRQLNLKRMVRPVRNIRNLTKADMKNNNATPKIDV
DPQTYEVFVDGNKITSEAATELPLTQRYFLF*

50

Sequence 119

Contig_0449_pos_2651_3103,

is similar to (with p-value 2.0e-48)

55 >sp:sp|Q07401|UREE_BACSB UREASE ACCESSORY PROTEIN UREE. >pir
:pir|D36950|D36950 ureE protein - Bacillus sp. (strain TB-90
) >gp:gp|D14439|BACUREA_4 Thermophilic Bacillus genes for ur
ease subunits and urease accessory proteins, complete cds. N
ID: g393296.

atgattatagaagaaattcaaggaaatattgctaatttatctcaagatgaaaagcaaaaa

catgtcgaaaaagtttatcttgaaaactcagatttggttaaactatataacgtgtttaa
 acagatcacggtaataagggatacgtcttaacaacctattgacctacaatatggt
 gatattttatatcaagacgatacaaacatgattattgtcgatgttaatagcgaagactta
 ttagttattaaacctagaaatttaaaggaaatgggagacattgctcatcaactaggtaat
 5 cgccatctgcctgcccaatttacagaaactgaaatgcttattcaatatgactatcttggt
 gaagattttattaaaagagttgggtatccctactcacatgaagacagaaaggtaaatcaa
 gcatttcgacatataggacattcacatgattga

Sequence 120

10 MIIEEIQGNIANLSQDEKQKHVEKVYLENSDLVKRIQRVKTDHGNEIGIRLKQPIDLQYG
 DILYQDDTNMIIVDVNSEDLLVIKPRNLKEMGDIAHQLGNRHLPAQFTETEMLIQDYLYV
 EDLLKELGIPYSHEDRKVNQAFRHIGHSHD*

Sequence 121

15 Contig_0449_pos_1467_1159,
 is similar to (with p-value 4.0e-37)
 >sp:sp|P42873|URE1_STAXY UREASE ALPHA SUBUNIT (EC 3.5.1.5) (UREA
 AMIDOHYDROLASE). >pir:pir|S38485|S38485 urease (EC 3.5.1.5) 62K chain -
 20 Staphylococcus xylosus >gp:gp|X74600|SXUREA BC_3 S.xylosus gene for ureA,
 ureB, and ureC genes for urease gamma, beta and alpha subunits. NID: g410513.
 atgccaaaggtcctggtgttacagtagtcgctttagcaccttcagaagcaccagttcctcc
 accgatatgcgttgtaataccactctcaagtgaacttcagcttggtcaggattgatgaa
 gtgcacgtgtgtatcgataccgccggcagtaacaattttaccttcagcagcaataatc
 25 agttgttgccacaaatgatgatgtcaacgttatccattatatctgggtttccagctttacc
 gatcttaaaaataataaccatttttaattccgatattctgcttaacaatcttgcataatc
 aataattaa

Sequence 122

30 MPRSWCYSSRFSTFRSTSSSTDMRCNTTLKCNFSLFRIDEVHVCIDTAGSNNFTFSSNNI
 SCCTNDVDVNIHYIWNSSFTDLKNITIFNSDICFNNLVIINN*

Sequence 123

35 Contig_0450_pos_6860_7486,
 is similar to (with p-value 8.0e-57)
 >nrl3d:pir||1GPHA Glutamine phosphoribosylpyrophosphate (prp
 p) Amidotransferase (EC 2.4.2.14), chain A - Bacillus subtilis
 is >nrl3d:pir||1GPHB Glutamine phosphoribosylpyrophosphate (prpp)
 Amidotransferase (EC 2.4.2.14), chain B - Bacillus subtilis
 40 >nrl3d:pir||1GPHC Glutamine phosphoribosylpyrophosphate (prpp)
 Amidotransferase (EC 2.4.2.14), chain C - Bacillus subtilis
 >nrl3d:pir||1GPHD Glutamine phosphoribosylpyrophosphate (prpp)
 Amidotransferase (EC 2.4.2.14), chain D - Bacillus subtilis
 45 atggtaataggcgtacctaattcatcattatctgcagcaagtgggttatgctgaagaaata
 ggctaccatatgaaatgggactagttaaaaatcaatatgttgctcgaactttatacaa
 cctactcaggaattaagagagcaaggtgtacgtgtgaaactgtcggctgttaaggatatt
 gttgatggtaagatatcgtaactttagatgattcgattgttcgaggtaacaacgattaaa
 cgcatagttaaaatgcttaaggattcaggagctaaccgcattcacgtaagaattgcttct
 50 cccgaattcatgttccctagtttttatggtattgacgatctacaacagctgaactcatc
 tcagcaagtaagtctcctgaggaaattaaaaatcatattggtgcagattctcttgcttat
 ttaagcgttgatggcttaatcgagtctataggacttgattatgatgcgccatatcatggc
 ttgtgtgtagaaagttttacaggtgattatccagcaggactttacgattatgagaaaaat
 tataaaaagcatttaagtgaacgtcaaaaatcatatatagctaataataaacattatctt
 55 gatagtgagggaatttacatgtctaa

Sequence 124

MVIGVPNSSLSAASGYAEEIGLPYEMGLVKNQYVARTFIQPTQELREQGVRVKLSAVKDI
 VDGGKDIVLVDDSIVRGTTIKRIVKMLKDSGANRIHVRIASPEFMFSPFYGIDVSTTAEI

SASKSPEEIKNHIGADSLAYLSVDGLIESIGLDYDAPYHGLCVESFTGDYPAGLYDYEKN
YKKHLSEKQKSYIANNKHVFDSEGNLHV*

Sequence 125

5 Contig_0450_pos_14672_15445,
putative peptide of unknown function
atggaccatagttccgcttcgaaaaaattaattaaagatatagagcaaaatcagtatgta
acagtaaaacatttatctcatgatgattttttatattgatgatttggtaaaaaagaagaa
gtcattgcaagccttgaaatacctaaggatttctcaaaacaccttaaagataatgattta
10 aataagactcttccattatatagcagagatgattttataggacatattgctatggaaata
atcagtcgatcattatacgaacaacaaatccctaataattattcatgaacatcttgatgat
atgaacaaccacaatccttagataaaagtgaacaatcttattattcgcttacacctcaa
tctaaaaataaaaagtgtagctatcaataaacaatgctcatcaatccatttcaattggcatt
gtatttgtgtcgctcatcttgtgaagtgttatccaaatcctattacatcaacgtcttaaa
15 cagaacgcacctctcgaaagatttatatttggtagcttatagtcacttaaaactatacttg
acttatatcagtgtagcatgtggtcatactgatgctcatgctattgatgatttagcctttta
atgcatcaaccattaagcattcttttctacttaaaaacactgatttatagttttattttat
gaggcaggtattgctttattacttttttaaaattaatgttcttagtcaccgtatattcatg
gctattatttatacggtagcgataggtatttatattttatggattcaattgtaa
20

Sequence 126

MDHSSASKKLIKIDIEQNQYVTVKHLSHDDFYIDDLVKKKEVIASLEIPKDFS KHLKDN DL
NKTLPLYSRDDFIGHIAMEIISRSLYEQQIPNIIEHLDDMKQPQSLDKVKQSYSLTPQ
SKIKSVAINKHAHQSSISIGIVFVVVIFVSVIQLLHQRKQNAPLERLYLVPYSQKLKLYL
25 TYISVHVHVLMLMLLMISLLMHQPLSILFYLKTLIIIVLFYEAGIALLLFKINVLSHRIFM
AIIYTVAGIILYLIWL*

Sequence 127

Contig_0450_pos_17961_18629,
30 is similar to (with p-value 3.0e-41)
>sp:sp|P40815|T3RE SALTY TYPE III RESTRICTION-MODIFICATION S
YSTEM STYLT I ENZYME RES (EC 3.1.21.5). >pir:pir|JN0658|JN065
8 restriction endonuclease (EC 3.1.-.-) - Salmonella typhimu
rium
35 atgcttaacatcatgacaacaaaaataatgattccaacgatatacaattttttatttatat
gcagactcatctttagattctggtatacctctaagcaatatgagtgaattaagttaaata
aatataataagagaatttaacaagcgttttgaagaaaaatatagtc aaagttaagaatat
aaaaaattagatttttctgctactacaacctttatgattcagaatatcagagtttaaa
gattgggtagatgcaaatattttaggtactaacgttgaaaataacattcaaaactgaaaaa
40 agatttttatatgaaagaccaccagtttagatatgatagtgtaaacacctgagttagagttg
ttaaaaaagaattacgataaaaaatgtaactgtatttggtaatttgcctaaaaaagcgata
caagttcctaaatatactggtggcactactacgcctgattttgtctatatgatagaaact
gatgaacaagatgcaaaataccttattgttgaaacaaaagcagaaaacatgagactagga
gataaaagtatttggtgaaatacaaaaaaattctttaacacattagataaattgaatatt
45 aaatatcaattagctactagcgcgcaagatgtttataatgaaattaaaaaattagatgat
tcaaaagtga

Sequence 128

MLNIMTTKIMIPTISNFLLYADSSLDSGIPLSNMSELSLNNIREFNKRFEKYSQSYEY
50 KKLDFSATTTIYDSEISEFKDWVDANYLGTVNENNIQTEKRELYERPPVRYDSVTPELEL
LKRNYDKNVTVFGLNPKKAIQVPKYTGTTTDFVYMIETDEQDAKYLIVETKAENMRLG
DKSIGEIQKFFNTLDNLNIKYQLATSAQDVYNEIKKLDDSK*

Sequence 129

55 Contig_0450_pos_18636_19928,
putative peptide of unknown function
atgatggggaaatcagaaaaaatttcattacttgaaaaagccaagatgggttagtagat
aaaaccggagcaaaacctaagttaccaacaactattaccgatttaaatcaagaacttta
gagggtatatagaataccactaaagtttctgtattataatgatagaaatggaagaattgct

tctgtaatatccagagtaagcgacgatataaaagttgcttatgaatttgaagataataat
 tataataaaagtattgaaaatatgatatatgaggctaatacttctgctttaaaaaacact
 aaaaaatctattaaagataaagggtcagcaagtatttggttatgtattagatgatggtaga
 gttattgacggaaatagaagggttactgcgcttagacagtagaacaagaaacaggaaat
 actttttattttgaagctggtattttaccatttacttatgataaaaagactgatcgagct
 5 aaaatcaagcaattagaacttgcaatacaaatgggtatagaaggaaagcaagattatgat
 aaagttgatgaagcgggtggatatttatcaacaattgaagttgaaaagttgatgactgta
 gcggactatgcaaagtgatcaaataaaacaaaaaaactattgaaaagcaattaggttct
 10 gcaaaaattaataagaaaatttttagattttattaatgctcaagagaattcttattatatt
 ataaaagatgctggcattttattcattatttgaagaagcagtagcctaagttagataaagct
 tatcctaaaggtggaccttcgtagaagatgctattgaaaaatttttagttttgctcct
 ttgcaaatccaatcagggacaagcacacgggcatatgctggaagagattattttgaaaat
 atcggtttttcaaagtgagggtagtcatttataacagaaacagaagatgctatcgat
 gggttgagagataaattagaagaaaaacgtgtagaatctaccgcagatttaaaagagtaca
 15 ctaagccaatcgatacctgaactacgagaagtaagttctcatataacaaagtgtgaagc
 aaaagtaaacgaaatgctaattgtggaagttttattgaaaatgttaaatcaatgtctgaa
 agtctcaatgatattgaaaaaggcaatgggttacctagtagtcttaattttgaacaattt
 aatctaaaacaattaaaagaaattagagaatgctaataagaataaataattgtagtaga
 gattgtagatatttatgaacatgaaatctga

20 Sequence 130

MMGKSEKISLLEKVQDGLVDKGTGAKPKLPTTITDLNQETLEVYRIPLKFLYYNDRNGRIA
 SVISRVSDDIKVAYEFEDNNYNKSIENMIYEANTSALKNTKKS IKDKGQQVFGYVLDDGR
 VIDGNRRFTALRQLEQETGNTFYFEAVILPFTYDKKTDRAKIKQLELAIQMIEGKQDYD
 25 KVDEAVDIYQTI EVEKLM TVADYANESNKTKTIEKQLGSAKLIRKFLDFIN AQENSYYI
 IKDAGIYSLFEEAVPKLDKAYPKGGPSLEDAIEKFFSFVLLQIQSGTSTRAYAGRDFYFEN
 IVFSNEGSHQFNTETEDAIDGLRDKLEEKRVESTADLKSTLSQSIPELREVSSSYNKVVS
 KSKRNNANVESFIENVKSMSESLNDEKGNLPSLNFQFNLKQLKEIREMLIRINNCSSR
 ELIDIYEHEI*

30 Sequence 131

Contig_0450_pos_22761_22258,
 putative peptide of unknown function
 atgaatacaatcaaaaagtagcgtacacacagaagcgatttttagcgatgatgaacaacac
 35 cgatacttactttaaaaagacgtggaatgaaaagaagcccacatgtacagtgataacgatg
 tccctcatttagacggcatattatcactcgatcttacaactgttcttatcctcaatcaa
 tttagcgaattcagaacgatacgggtgctgtatatttagtgaatttttttgcgaattataaa
 accccagataatctcaaacatattaaagaccttatgataaacatacagacagacactta
 atgaaagcaaatgagagtgacacagtaattctagcttatggagcctatggaagcga
 40 ccatttggttatcgaacgtgttgaacaagtgtggaatgttgaagcctcacaaaaagaaa
 attaaaaagctcataaaccagcaacaaatgaaatcatgacccactcaatcctaaagca
 cgccaaaaatggacattgaaataa

45 Sequence 132

MNTIKSTIHTAIFSDDEQHRYLLKKTWNEKKPTCTVITMSPHLDGILSLDLTTVLILNQ
 LANSERYGAVYLVNLFNLIKTPDNLKHIKEPYDKHTDRHLMKAISESDTVILAYGAYAKR
 PFVIERVEQVMEMLKPHKKKIKKLINPATNEIMHPLNPKARQKWTLK*

50 Sequence 133

Contig_0450_pos_17953_17219,
 putative peptide of unknown function
 gtgattacgcaaggagatagaattggatggttaaactcctcttatattgatattaattgct
 atattcattgtgacattaattgcatttttatattttgaaaaacgtcaagatgaacctttt
 atagattttaagttttttcaaataatgttttatattggaacaacattagccaacttgatg
 55 gtgaacatggatattgggttcatttagcattatttaatatatttatgttcaagacgataaacat
 ctatcagctgcacaagccggtttaattacaattccatatatgctgtgtagtttgtaagt
 attcgtgttggtgaacgttttatgcaaaaaagaggaccgcaattgccattgatgttaggt
 ccggtatcaattactgttggtattatacttttagcattcacttctttgcctaataatgatt
 tatttatattgtggcatgtattggctttatctttataggcttaggattaggatttttgct

acacccgcgctatctacagctgtatctaattgtccagctgaaaaagcaggtactgcatca
 ggaattatcaaaatgacttctacactaggtgcagcatttggaatcgctgtgtgacaaca
 atatatacggcattatctgtaaatcaccggcatatttagcagctactatcgcatctatc
 gtgggtgcaggttttagtgtttatcgcatctattgctggcggtattgtttaattcctaaaaag
 aatgtagatatctaa

Sequence 134

VITQGDRIGWLNPLILILIAIFIVTLIAFYIFEKRQDEPFIDLSLFSNNVYIGTTLANLM
 VNMDIGSLALFNIVYQDDKHLASAAQAGLITIPYMLCSLLMIRVGERFMQKRGPLPLMLG
 PVSITVGIILLAFSLPNMIYYIVACIGFIFIGLGLGFFATPALSTAVSNVPAEKAGTAS
 GIIKMTSTLGAAFGIAVVTIYTALSVNHPAYLAATIAFIVGAGLVFIAFIAAYCLIPKK
 NVDI*

Sequence 135

Contig_0450_pos_11730_11380,

putative peptide of unknown function

atgattggtatgtcggttaatacaactaggtgcttttaagaagctttaccatttttaatg
 actgcagctgaaatggacgatgatagagatttagaagtacagtttcaatatgggttagta
 ctatgccaaactcgaaatgtttgatgaagctattaaacaattaaataagggtccttctatc
 gattcacagcacgtagatggatatataatcttggttagcaacatatatgaaaaatgaa
 aatttagatgaagcaattgcatatttgacaagcaatatcaattgatgaaaaacattta
 cttagtcaacatgcattaaagacattcaaaacaatgaaagaggaggaataa

Sequence 136

MIGMSFNQLGAFKEALPFLMTAAEMDDDRDLEVQFQYGLVLCQLEMFDIAIKQLNKVLSI
 DSQHVVDGIYNLGLATYMKNNENLDEAIAYFEQAISIDEKHLLSQHALKTFKTMKEEE*

Sequence 137

Contig_0450_pos_11378_9129,

putative peptide of unknown function

atgtctgaccctacactttttgattattcaatgatcaaaggtacagttgatgctatttta
 tttcaaaatacggataattttttatactgttctaaaagtagatactatagaatcaaatgaa
 aaatttgatagtagtgccaactgtggttaggggttcttcccaatgtagttgaaggcgatgtt
 tatacttttaaaagggaagtcgtacaacatccacgttatggtaagcaattaaagggtgaa
 acatttgaaaaagaattacctcaaactaaagaagccattattagttacttatcaagtgat
 ttatttaaggcatcggtaaaaaacggctcaaaacattgtaaatacactaggtgaaaat
 gctataaatgatattttaactcgtccagaaatcttagaaagtgtacctagtttaccaaag
 aagaaacaaaagcaaattgctgatcagattaatgcaaaccaagaatctgagaaaattatg
 atacgtttacacgacctagggtttggtccgaaattatcaatggctatatatcagttctat
 atgggtgatacttttaaatgtcttagataaaaaatccttaccatttagtatatgacattaaa
 ggtattgggttttaataaggctgaccaacttgctcgaaatgtcggtattgagccacattca
 cctgaaagattaaaagcagcattattatttacgttagaagaagaatgtatcaaacaagga
 catacatatctacctcgtacaattgttatagaaacaacacaaaatttactcaatgaagat
 attgagaaaccaattgaaacagagcaattactagaaatcattgacgttttatcagaagag
 aaaaaattaatatctgaagctgatcaggtatcaattccaagttttatactattcagaattg
 aaaagtggtgcaaaaacttataccgaattaaaacaaacacatctaaattaaaagaaatagaa
 cagtcgtatttcaaaatacatattggtgatattgagtcacaaaatgagggttaattactct
 gcctctcaaaaagaagcgcttgaaacagcaataaattctaaaattatgcttttaactggt
 ggtccgggtaccggtaaaaccacagtcattaaaggtatagttgaatttatgcagaaata
 catgggctctcgtcgattatgatgattacaatgaagatgattatccagtagtggttagct
 gacccactggctgcttctaagcgcttcacgaatcgacaggtttagaagcaatgaca
 attcatcggttaaatcggttggaaccaagatacacacaccacaggatattttagaaaatgag
 atcaatgcaagactcattatcatcgatgaaatgtcaatggtagatacttggttggtccat
 caatttttaagcgctgtgcttttagaagcacaaaattgtatttgctggagatgaagatcag
 ttaccatcagtaggtccaggacaggtatttaagaccttattgattctgaaataataaccg
 cgtgttaattctaccgaagtatatcgtcagcaagatgggtccagtagttattgacttagct
 caccgtatgaaattaaatgaacctatcgatattactaaacgttatcatgatcgtagtttt

attcgttgtggtacgaatcaaattccagacggttgttgataaagtagttaaaagcgctgta
 gctaaaggctatgatatgagtgatatacaagtttggctcctatgtataaaggtaacgct
 ggtattaagagacttaaccaagttctacaatctattcttaatccgaagcaacaagatgat
 cgtgaaatagaatttgggtgaagctgtgttttagaaaaggggataaagtacttcagttagtt
 5 aatcgacctaataatgataatatatttaattggggatataaggtataatagtaggtatattttgg
 gccaaagaaaatgctctaaataaggatgtgttagttgtagattttgaaggtaatgaaat
 acatttactaaacaagatttaattggaactaacacatgcatattgtacatctatccataaa
 tcacaagggttcagaatttcctattgttaattatgcctattgttagacaatattataggatg
 ttacaacgtcccattctttatacaggattaactagagctaaacaatcacttgttttatcg
 10 ttaaagagagatatacattttttatttttttaagatttctcagaaaaattaggttttct
 tcattttaatttttaattcttagccatttataa

15 Sequence 138

MSDPTLFDYSMIKGTVDAILFQNTDNFYTVLKVDITIESNEKFDSMPTVVGFLPNVVEGDV
 YTFKGQVVQHPRYGKQLKAETFEKELPQTKEAIIISYSSDLFGIGKKTQNIIVNTLGEN
 AINDILTRPEILESVPSPKPKKQKQIADQINANQSEKIMIRLHDLGFGPKLSMAIYQFY
 MGDTLNVLDKNPYQLVYDIKIGIFNKADQLARNVGIIEPHSPERLKAALLFTLEECIKQG
 20 HTYLPRTIVTETQNLNEDIEKPIETEQLLEIIDVLSEKKLISEADQVSIPLSYSEL
 KSVONLYRIKTNTSKLKEIEQSDLIHIGDIESQNEVNYSASQKEALETAINSKIMLLTG
 GPGTGKTTVIKIGIVELYAEIHGLSLDYDDYNEDDYPVVLAAAPTGRASKRLHESTGLEAMT
 IHRIGWNQDTQPQDILENEINARLIIIDEMSMVDTWLFHQFLSAVPLEAQIVFVGDEDQ
 LPSVGPQVFKDLIDSEIIPRVNLTEVYRQDQSSIIDLAHRMKLNEPIDITKRYHDSF
 25 IRCGTNQIPDVVDKVVKSAAVAKGYDMSDIQVLAPMYKGNAGIKRLNQVLQSILNPKQDD
 REIEFGEAVFRKDKVLQLVNRPNDNIFNGDIGIIVGIFWAKENALNKDVLVDFEGNEI
 TFTKQDLMEALTHAYCTSIHKSQGSFPIVIMPIVRQYYRMLQRPILYTLTRAKQSLVLS
 LKRDIHFYFLFLRFLRKIRFFSFNFNLSHL*

30 Sequence 139

Contig_0450_pos_5605_5021,
 is similar to (with p-value 8.0e-52)
 >sp:P54378|GCST_BACSU PROBABLE AMINOMETHYLTRANSFERASE (EC
 2.1.2.10) (GLYCINE CLEAVAGE SYSTEM T PROTEIN). >gp:gp|D8443
 35 2|BACJH642_194 Bacillus subtilis DNA, 283 Kb region containi
 ng skin element. NID: g2627063. >gp:gp|Z99116|BSUB0013_168 B
 acillus subtilis complete genome (section 13 of 21): from 23
 95261 to 2613730. NID: g2634723.
 atggcaatgtttgaattcaaacagaaacgtacaaatcttttggtaaatctattattctttcg
 40 cagctctggttatactggagaagatggctttgaaatttactgtaagcaagaagataactaag
 gatataatgggagcaattattagaatacagatgttacaccatgcggttttagtgctcgtgat
 acgctaagacttgaagcaggattacctttacatggtcaagatttatctgaatcaattact
 ccttatgaaggaggatagccttcgctgctaaaccgttaattgaaaatcattttattggc
 aaatccgtactcaaagctcaaaaagaaaatggttccgagcgtagaacagtaggtcttgaa
 45 ctattaggtaaaggcattgctagaacagggttatgacgtactagatgaaaatagtaatgaa
 attggtttcgttacatcaggaacacaatccccatcttctggtaaatctatagcacttgca
 ataatagatagagatgcatttgaaatgggcaaaaaagtaattgtgcaaatacgtaagcgt
 caagttgaggcaaaaatagttaaaaaaaatcaaattgagaaataa

50 Sequence 140

MAMFEFKQNVQIFGKSIILSQSGYTGEDGFEIYCKQEDTKDIWEQLLEYDVTPCGLGARD
 TLRLEAGLPLHGQDLSESITPYEGGIAFAAKPLIENHFIGKSVLKAQKENGSRRTVGLE
 LLGKGIARTGYDVLDENSNEIGFVTSQTSPSSGKSIALAIIDRAFEMGKKVIVQIRKR
 QVEAKIVKKNQIEK*

55

Sequence 141

Contig_0450_pos_4620_3655,
 is similar to (with p-value 3.0e-97)
 >sp:sp|P54376|GCS1_BACSU PROBABLE GLYCINE DEHYDROGENASE (DEC

ARBOXYLATING) SUBUNIT 1 (EC 1.4.4.2) (GLYCINE DECARBOXYLASE)
 (GLYCINE CLEAVAGE SYSTEM P- PROTEIN). >gp:gp|D84432|BACJH64
 2_195 Bacillus subtilis DNA, 283 Kb region containing skin e
 lement. NID: g2627063. >gp:gp|Z99116|BSUB0013_167 Bacillus s
 5 ubtilis complete genome (section 13 of 21): from 2395261 to
 2613730. NID: g2634723.
 atggatgtagcaaatcttctatgtatgatggatgactagttttgctgaagcatgtata
 ttggcactaagtcatacgaaaaaaataaaattgtagtttcaagtgactacattatcaa
 gctttacaaattctacacacatacgccaaaactcgtgatgaatttgaaataattgaagtt
 10 gatcttaaagggtactattactgatttagagaaattagaacaacttatcgatgacaacaca
 gcagctgtcgtgtccaatatcccaatttttatgggtctattgaagatttagaacaatt
 aataactatataaaggataaaaaagctttatttatcgtatatgccaatccactttcttta
 ggattactaacaccccaggtacattcggggcagacatagtagtgggagatacacagcct
 tttgggtattcctacacaatttgggggtccgcattgtggatactttgctacaacaaagaaa
 15 ttaatgagaaaagtacctggtcgattagttgggcaaactcaagatgacgaaggtaatcgt
 ggatttgttctcacgttacaagctagagaacaacatatccgccgtgataaagcaacttct
 aatatttgttcaaataagcttttaaatgcacttgcattctcaatagcaatgtcagcttta
 ggtaaacaaaggattttatgaaattgcagttcaaaatcttaaaaatgccaatatgccaaa
 aataagtttgaagaacatggttttgaggtactaaaagcacaatcttttaaatgaatttga
 20 gtcaaatattaatcaaccaataaaaaataataatcttaaatagcagaatatggatatatt
 ggtgggtttgacttaggtgaagtatctgatgatttttaaaaaccatatgttagtagcagtt
 acagagttaagatctaaagatgaaatcgatgatttcgttacgaaagcaggtgagttaaat
 gattag

25 Sequence 142
 MDVANSSMYDGMTSFAEACILALSHTKKNKIVVSSGLHYQALQILHTYAKTRDEFEIIIEV
 DLKGTITDLEKLEQLIDDNTAAVAVQYPNFYGSIEDLEQINNYIKDKKALFIVYANPLSL
 GLLTPPGTFGADIVVGDTPFGIPTQFGGPHCGYFATTKLMRKVPGRLVGQTQDDEGNR
 GFVLTLQAREQHRRDKATSNICSNQALNALASSIAMSALGKQGIYEIAVQNLKNANYAK
 30 NKFEHGHFEVLKAQSFNEFVVKFNQPIKNINLKLAEYGYIGGFDLGEVSDDFKNHMLVAV
 TELRSKDEIDDFVTKAGELND*

Sequence 143
 Contig_0450_pos_3419_2154,
 35 is similar to (with p-value 0.0e+00)
 >sp:sp|P54377|GCS2_BACSU PROBABLE GLYCINE DEHYDROGENASE (DEC
 ARBOXYLATING) SUBUNIT 2 (EC 1.4.4.2) (GLYCINE DECARBOXYLASE)
 (GLYCINE CLEAVAGE SYSTEM P- PROTEIN). >gp:gp|D84432|BACJH64
 2_196 Bacillus subtilis DNA, 283 Kb region containing skin e
 40 lement. NID: g2627063. >gp:gp|Z99116|BSUB0013_166 Bacillus s
 ubtilis complete genome (section 13 of 21): from 2395261 to
 2613730. NID: g2634723.
 atgaaatataatcctaaaatcaatgaaaaggtagcgctatttctggttttagtgaatct
 catcctttacaagaagaagaacacgttcaagggttctctgaaattatatatagtttaca
 45 gaagaattgaaggaaattactggatggatgaagttaccctacaacctgctgcaggtgca
 catggtgagtgactgctttaatgattttcaaagcttatcatgaaaaaatggacaaagc
 catcgtgatgaagtaatagtgctgattcagcacatggtaactaatcctgcttctgcctca
 tttgctggatttaaatcagtaactgtaaaatctaatacaacgtggggaagttgacatagaa
 gatttaaaaagagtagtaaacgataatacagctgcaatcatgttaactaatccaaatata
 50 ttaggtatatttgaacaggatattattgaaatagggaaaatcgttcatgaagcaggaggt
 ttattatattacgatggagcaaatttaaatgctattttagataaagtacgtcctgggtgat
 atgggctttgatgcggtacatcttaatttgcacaaaacattcactggtcctcatggcggt
 ggtggaccaggatcaggaccagttggagtagtagagaaattagccagttatctacctaag
 cctatggttataaaaagataacgataggtataaatatgataatgatattccaaattcaatt
 55 ggacgagtaaaaaccgtttttatggaaaatttcggcatttttaagagcatatacttatatc
 agatcaatgggagccaatggtttaaaaagaagtatctgaagctgccgttcttaatgcgaat
 tatataaaatctcgcttataaaatcactttgaaattccggttcaatcaatattgtaaacat
 gaatttgtattaaagtggaactttacaaaaacaatatggtgtcagaacattagatatggct
 aagcgactgttagattttggtgtgcatccacctacaatatattttcctctcaatgtcgaa

gaaggaatgatgattgagccaacagaaactgaatctaaagaaacacttgattactttatt
 gatgcatgattcgaatcgctgacgaaacaaaaaatgatccagataaagttttagaagca
 ccacatacgactataattgatcgattagatgagaccactgcagcacgaaaaccaattcct
 aattttgaagaacttaaggacgaaaagtataaagaacacacaaatattgattctgaagat
 5 aattaa

Sequence 144

MKYNPKINEKVARISGFSESHPLQEEHVQGSLEIIYSLQEELKEITGMDEVTLQPAAGA
 HGEWTALMIFKAYHEKNGQSHRDEVIVPDSAHGTPASASFAGFKSVTVKSNQRGEVDIE
 10 DLKRVVNDNTAAIMLTNPNTLGIFEQDIIIEIGKIVHEAGGLLYYDGANLNAILDKVRPGD
 MGFDAVHLNLHKTFTGPHGGGGPGSGPVGVVEKLASYLPKPMVIKDNDRYKYDNDIPNSI
 GRVKPFYGNFGIYLRAYTYIRSMGANGLKEVSEAAVLNANYIKSRLKNHFEIPFNQYCKH
 EFVLSGTLQKQYGVRTLDMAKRLLDGVPPTIYFPLNVEEGMMIEPTETESKETLDYFI
 DAMIQIADETKNDPKVLEAPHTTIIIDRLDETTAARKPILKFEELKDEKYKEHTNIDSED
 15 N*

Sequence 145

Contig_0450_pos_0_722,

is similar to (with p-value 2.0e-53)

20 >sp:sp|P54511|YQHM_BACSU HYPOTHETICAL 22.8 KD PROTEIN IN GCV
 T-SPOIIIAA INTERGENIC REGION. >gp:gp|D84432|BACJH642_198 Bac
 illus subtilis DNA, 283 Kb region containing skin element. N
 ID: g2627063.

atgattatgactgaaatatggaattttataaatactggaagcaaaaatccttattataat
 25 atggcaatggacgaagcggtactaaattttgtatcgctggagaaatcgatccagttata
 agattttatacttggaatcctgcaacactctcaataggctactttcagcgtctccaaaaa
 gaaattgatattgataaaagtaaaagaaaagggtatggcttagtaagacgtcaaacgggt
 ggtagaggcgtgttacacgataaagaattaacatatagcgttattgttcttgagttctcat
 ccaaatatgccttcaactgtaactgaagcttataaaattatttcacaaggattattagaa
 30 ggttttaaaaatttaggttttgaaacttatttcgctatccccggttctaaagaagaacga
 gacaaattaaagcaaccacgaagttcagtatgttttgatgcacctagttggtatgagctt
 gtatgtagaaggcagaaaaattgcaggtagcgtcaaaccagacaaaaagggtgtcattctt
 caacatggttcaattttacaagatatagatatcgatgatttatttgatatgtttaaattt
 aaaaatgaacgactaaaagcaaaaatgaaagaaaattttgttcaaaaagctgtagctatt
 35 aatgacatttcaaatcaacatattacattaaatgaaatggagaacgcctttgaggcaggt
 tt

Sequence 146

40 MIMTEIWNFINTEGSKNPYYNMAMDEALLNFVSRGEIDPVIRFYTNPATLSIGYFQRLQK
 EIDIDKVKEKGYGLVRRQTGGRGVLHDKELTYSVIVPESHNPMPSTVTEAYKIIISQGLLE
 GFKNLGFETYFAIPRSKEERDKLKQPRSSVCFDAPSWYELVVEGRKIAGSAQTRQKGVIL
 QHGSILQDIDIDDLFDMFKFKNERLKAAMKENFVQKAVAINDISNQHITLNEMENAFEAG
 X

45 Sequence 147

Contig_0451_pos_2108_3121,

putative peptide of unknown function

atggaacgattttgtgtgtgtaaatacaatttaactatattcaaatgaatccggttagaagcc
 aaattttaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagcctt
 50 gcttggttaggaccgacgctcaaacaaacaggcagcttacctatacatgagttaatattc
 tttgaattaagagaacgtgccggttttcatctagaaatcgagaatgaacaaaatcgactt
 aaatttcagattcttgaattactccatcaaacattccctgggttagaaagattatttagt
 agtcgatattcaatcattgcactcaacatcgagaaatttttactcatccagacgtggtt
 cttgatatcgacaaggatgacttattacacatatattcaattctacagataagggaatg
 55 tcaatggataaagctacaaaatatgcacttcaattaagagtgattgctcaagaaagctat
 cctaattgctgatagacattcctttctagtcgaaaaattacgcttacttattcaacaatta
 aaacaatctattcatcatctcaacaatttagatgatgccatgattcaattagcacaacaa
 ctcgattattttgaaaatattcattcgatacctgggtattggtaagctaagcacagctatg
 attattggggagattggtgatattaagcgatttaaatcaataaacaactcaacgctttt

gtaggcattgatatcaaacgatatcaatcaggtcatacacactgtagagataccatcaac
 aagcgtggtaataaaaaagcgagaaaacttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgtcgattattactacaaactaagaaagcagcct
 aatgagaaacctcataagactgccatcattgcttgataaatcgattattaaaaacgatt
 5 cattatctggtaatgaatcataaattgtacgattatcaaatgtcaccacattag

Sequence 148

10 MERFCCVNQINYIQMNPLEAKFKTSALRSWKTDQADAHKLACLGPTLKQTGSLPIHELIF
 FELRERARFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSIIALNIAEIFTHPDVV
 LDIDKDVLIITHIFNSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLRLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLSTAMIIGEIGDIKRFSKNQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLEFWVIMNIIRGQHHDNHVVDYKYKLRKQP
 15 NEKPHKTAIIACINRLKTIHYLVMNHKLYDYQMSPH*

Sequence 149

Contig_0451_pos_4254_4778,

putative peptide of unknown function

20 atgcttgatacatctgacatcgaaagggaaaaacgatttttagatgtgggatgtaatacaaggc
 ggatttttacgacagttatacgatacaaacaccgtttaaaaaaggtgttgcatagattta
 gcacgtttatctttggaaaaggcagagacattaaaaggacaacgtccacttacatactat
 ttaacagataaaccgcaagaaacgaagcacgtgtttgatacggcagtaagtacgtctgtc
 ttgtacttaatagaagatattccgcaacatgcaaaaagatttaaaaagaggtattgaaacca
 25 ggcggtgtttattacgcttcattcgcggttttaactaataacccaagtcgtcagtttatg
 gatgacacgattaatcaatatggtgcaacaccttctcagaatcactctctctaaacatac
 gttgatagctttgtggatgcaggatttgaaagttgcagtaatgaaaaagcatgtacttata
 aatatttatctaaaaagattggatcaatcgttgaaacaacttta

Sequence 150

30 MLDTSIEGKTILDVGCNQGGFLRQLYDTPFKKGVGIDLARLSLEKAETLKGQRPLTTY
 LTDKPQETKHVFDTA VSTSVLYLIEDIPQHAKDLKEVLKPGGVYASFADLTNNPSRQFM
 DDTINQYGATPSQNHSLKHIVDSFVDAGFEVAVMKHVLINIYLRKLDQSLKQL*

Sequence 151

Contig_0451_pos_5060_5626,

is similar to (with p-value 1.0e-19)

>sp:sp|P23477|ADDB_BACSU ATP-DEPENDENT NUCLEASE SUBUNIT B. >
 pir:pir|A39432|A39432 ATP-dependent exonuclease synthesis pr
 40 otein AddB - Bacillus subtilis >gp:gp|M63489|BACADDAA_1 Baci
 llus subtilis ATP-dependent nuclease (addA) and (addB), and
 open reading frame 3, partial cds. NID: g142438. >gp:gp|Z991
 09|BSUB0006_138 Bacillus subtilis complete genome (section 6
 of 21): from 999501 to 1209940. NID: g2633260. >gp:gp|Y1408
 45 1|BSY14081_20 Bacillus subtilis chromosomal DNA, region 92 d
 egrees: region between comK and addAB. NID: g2226171.

atggatattgtattacaaaacaaggagcgttttaggtcttacagatattgtgaaaccaggg
 ggtctacttttatttccatgtccatgaaccgcgtattaaatttaaaagttgggcagatata
 gatgaagaccaatttcaaaaagactatatcaaaaactttaaaatgagtggtttgctta
 50 cgtgaccaagaagtgttagacgcttttagatattagacttgaaccaaagtataattcgat
 attgtccaatagcattaacagctaaaggcgctataaatcaacgtagtagtaagtagct
 gatgaaaacatcatttatcaattaatagaacataataagaagaattttatcgagacagcc
 agccacattatggatggacatacggaaagtgccacccttgaagtacaaacaagtattacct
 tgtcaattttgttaattataaatcagtttgcgtgtagacggattaatagatagtaagcgt
 55 tatagaacagtagaatgaatgataaaaccattagatttaattcaacaattaagaaatgaa
 ggtggtgaaagacatgattccaactaa

Sequence 152

MDIVLQNKERLGLTDIVKPGGLLYFHVHEPRIKFKSWADIDEDQFQKDYIKNFKMSGLLN

RDQEVLDALDIRLEPKYNSDIVPIALTAKGAINQRSSKVADENIIYQLIEHNKKNFIETA
SHIMDGHTEVAPLKYQVLPQFCNYKSVCHVDGLIDSKRYRTVDESIKPLDLIQQLRNE
GGERHDSN*

5

Sequence 153

Contig_0451_pos_6249_9269,

is similar to (with p-value 0.0e+00)

10 >sp:sp|P23478|ADDA_BACSU ATP-DEPENDENT NUCLEASE SUBUNIT A. >
pir:pir|B39432|B39432 ATP-dependent exonuclease synthesis pr
oteins AddA - Bacillus subtilis >gp:gp|M63489|BACADDAA_2 Baci
llus subtilis ATP-dependent nuclease (addA) and (addB), and
open reading frame 3, partial cds. NID: g142438. >gp:gp|Z991
15 09|BSUB0006_139 Bacillus subtilis complete genome (section 6
of 21): from 999501 to 1209940. NID: g2633260. >gp:gp|Y1408
1|BSY14081_21 Bacillus subtilis chromosomal DNA, region 92 d
egrees: region between comK and addAB. NID: g2226171.
atgcagcttatcaatgatttagcaatgatttttatgaaagcaggatagaggaattacaa
20 aaaagttatgacttattctcaatgatggaaagtgttgataagcagcttgaagtattgaa
accgaacgcattgtttattactaaagctattgaaggtaaagtattaaatacagatgttattc
acgcaacatgaatttatgagtcgttttccggcaataaataagcaagataaaaagaagcaaat
gaaggcatggaagatgctttaaatgaagcaaaacaacattatgataaatataaatcttta
gttatgaaagtaaagaatgattattttctagaaatgcagaagatttgcaaaagagatattg
25 caacaactcgcacctcgagtggttatttagctcaaatagttcaagatgtgattcaatca
tttgggtgttcaaaaacgaagtcgtaataattttggatttttcagattatgaacattttgca
ttacgcattcttactaacgaagatggctcaccttcgcgtatcgctgaaacgtatcggtgaa
cattttaaagaaatcctagttgatgagatcaagataactaatagagtgcaagaaaaata
ttatcttgtattaaaactgggtgaagaacacgatggtaactgttcatgggtggggatgtg
30 aagcagctctatttataaaatttagacaagctgacctagttatttattgaaaaatataat
cgcttttctagtagtggaatgaaagtggcttgccgattgacttatcgcaaaactttcgt
tcgagacaggaagtgttatctacaaccaattacttgttcaaacatatgatggatgaacaa
gtaggagaaatttcatatgatgatgcagcgcaattgtattttgggtgcaccatatgacgaa
gtttcacatcctgttcaattacgagcacttgttgaggcaagttcagaaaaatagtgactta
35 actggaagtgaacaagaagcgaattacattgttgaacaagttaaagatattattaatcat
caaaacgtatacgatatgaaaacaggtcaatacagaaaagcaacatataaagatatcgta
attttagagcgaagttttgggtcaagcgctaatcttcaacaagcttttaaaaataatgat
atcccttttccacgtaaatagtaaggaagggtattttgagcaaaactgaagtacgtcttgtg
ctttcatttttaagaacaatagataatccacttcaagacattttatttagtgggattgatg
40 cgttctgtaatatatcaatttactgaagaagaattagctgaataaagagttgtaagccct
catgatgattatttatacaattctataaaaaattatagttgatgaaaaagctgattct
agattgggtgacaagttaaatacgttttattcaggatatacaaaaatatcaaaattatagt
ctaagtcaaccgggtttaccaattaattgataaattttataatgatcattttgtaattcag
tacttttagcggcttatttgagggtaaaggtagaagagcaaatctgtatgggctatttaaat
45 aaagctgttgaaatttgaaaattcaagtttcagagggtttattccaatttatcgttttatt
gatgagcttattgatcgtaaaaaagattttgggtgaagaaaatgtcgtaggtcctaacgat
aatgtggttagaatgatgacgattcacagtagtaaaaggattagaatttccatttgtaatt
tactcaggattatctaaaaaattcaacaagggtgacctgaatgcaccagttatttctaaat
caacaatatggttttaggtatggattattttgatgtaataaagatatggcttttcttca
50 cttgcctctgtggcatatagagcaataaatgaaaaagaacttatatcagaagagatgcgt
ttaatctatgttgcttgacacgagcaaaagagcaacttattttagttggaagagtcaaa
gatgaaaagtgcgttaattaaatatgaacaattagctgtttcagacacacatatagcagtt
aatgaacgccttactgctaccaatccatttgttctaatttatgggtgttttggttaagcat
caatcgcttcttattgccaatgatcaaagatttgaaagagatattgatcaattaaattct
55 gaagtgaagccacgtgtatcaatagtgattgatcattatgaggatgtttcaactgaagaa
gtagtcaatgataatgaaataagaacaatcgaagaattaaaggccataaatactggtaaat
gaagatgtgaaaattaaaattcatcaacagctttcttatgactatccttttaagttaac
acgatgaaaccatctaaacagtcggtatcagagttaaaacgtcaattagaactgaagaa
agtaatacaaaattatgatagagtacgtcaatatcgatttgggtgttgcatcatatgaaaga

5 cccaagtttcttaccacaaacaaaaaaagaaaagcaaataagaggactttaatgcat
 acagtcagtgcaacacttaccttttagagaacaacgtttaacaaaagacgaattatttcaa
 tatatcgatcgattgattgacaaacaacttattgatgaagatgcaaaagaggatattaga
 atagatgagattatgcatttcattgatggccctctctatatggaaatagctcaagctgac
 10 aatgtttatactgaattaccttttgggtaaatcaaattaaagttgatggacttacaagt
 gaagatgaagatgtatccattattcaaggtatgattgatttaatatatgaaagtgacgga
 caattttactttgttgattacaaaacagatgcttttaatagaagaaaaggtatgagtgat
 gaagaaatagggaaatcagctcaaagaaaaatatcagatacaaatgacgtattatcgaaat
 acttttagaaaccatactttaaacgacccgtaaagggttacttatattttttcaaatttggg
 acattagaaatagatgattaa

Sequence 154

15 MQLINDLAMIFMKAGYEELQKSYDLFSMMESVDKQLEVIETERMFITKAIEGKVLNTDVI
 TQHEFMSRFPAINSKIKEANEGMEDALNEAKQHYDKYKSLVMKVKNDFSRNAEDLQRDM
 QQLAPRVAYLAQIVQDVIQSFVQKRSRNILDFSDYEHFALRILTNEGSPSRIAETIRE
 HFKEILVDEYQDTRNVQEKILSCIKTGEEHDGNLFMVGDKQSIYKFRQADPSLFIEKYN
 RFSSSGNESGLRIDLSQNFRRSQEVLSTTNYLFKHMMEQVGEISYDDAAQLYFGAPYDE
 VSHPVQLRALVEASSENSDLTGSEQANYIVEQVKDIINHQNVDYDMKTGQYRKATYKDVI
 20 ILERSFGQARNLQQAFFKNNDIPFHVNSKEGYFEQTEVRLVLSFLRTIDNPLQDIYLVGLM
 RSVIYQFTEELAEIRVVSPhDDYFYQSIKNYMIDEKADSRVLVDKLNRFIQDIQKYQNY
 LSQPVYQLIDKFYNDFHFIQYFSGLIGGKRRANLYGLFNKAVEFENSFRGLFQFIRFI
 DELIDRKKDFGEENVVGPNDNVVRMMTIHSSKGLEFPFVIYSGLSKKFNKGDLNAPVILN
 QOYGLGMDYFDVNKDMAFPSLASVAYRAINEKELISEEMRLIYVALTRAKEQLILVGRVK
 25 DEKSLIKYEQLAVSDTHIAVNERLTATNPFVLIYGLAKHQSPSLPNDQRFERDIDQLNS
 EVKPRVSIVIDHYEDVSTEEVNDNEIRTIIEELKAINTEGNEVDKIKIHQQLSYDYPFKVN
 TMKPSKQSVSELKRQLETEESNTNYDRVRQYRIGVASYERPKFLTQTKRKRKANEIGTLMH
 TVMQHLPFREQLRKDELQYIDRLIDKQLIDEDAKEDIRIDEIMHFIDGPLYMEIAQAD
 NVYTELPFVFNQIKVDGLTSEDEDVSIQGMIDLIYESDQGFYFVDYKTDFAFNRRKGMDS
 30 EEIGNQLKEYQIQMTYYRNTLETILKRPVKGYLYFFKFGTLEIDD*

Sequence 155

Contig_0451_pos_11640_12653,

putative peptide of unknown function

35 atggaacgattttgttggtgtaaatcaaattaactatattcaaataaatccgtagaagcc
 aaattttaaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagcct
 gcttggttaggaccgacgcttaaacacacagacaacttacctatacatgagttaatattc
 tttgaattaagagaaacgctccgttttcatctagaaatcgagaatgaacaaaatcgactt
 aaatttcagatccctgaattactccatcaaacattccctgggttagaaaagattggttagt
 40 agtcgatattcaatcattgcactcaacatcgagaaatctttactcatccagacatgggt
 cttgatatcgacaaggagggtactgattacacatatattcaattctacagataagggaatg
 tcaatggataaagctacaaaatatgcacttcaattaagggtgattgctcaagaaagctat
 cctaattgctgatagacattcctttctagtcgaaaaattacgcttacttattcaacaatta
 aaacaatctattcatcatctcaacaattagatgatgccatgattcaattagcacacaaca
 45 ctcgattattttgaaaatattcattcgatacctgggtattggtaagctaagcacagctatg
 attattggggagattgggtgatattaagcgattttaatacaataaacaactcaatgctttt
 gttggcattgatatacaacgatatacaatcaggtcatacacactgtagagataccatcaac
 aagcgtggtaataaaaaagcgagaaaacttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgatgattattactacaaactaagaaagcgcct
 50 aatgagaaacctcataagactgccatcattgcttgataaatcgattattaaaaacaatt
 cattatcttgtaataatcataaattgtacgattatcaaattgtcaccacattag

Sequence 156

55 MERFCCVNQINYIQMNPLEAKFKTSALRSWKTQDAHKLACLGPTLKQTDNLPIHELIF
 FELRERVFRHLEIENEQNRLKFQILELLHQTFPGLERLFSSRSYIIALNIAEIFTHPDMV
 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLRLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFKSNKQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLFWVIMNIIRGQHHYDNHVVDDYKLRKQP
 NEKPHKTAIIACINRLKTIHYLMNHKLYDYQMSPH*

Sequence 157

Contig_0451_pos_15717_16034,

5 is similar to (with p-value 1.0e-38)

>gp:gp|Z79580|BS168NPRB_7 B.subtilis nprB gene. NID: g162092

1. >gp:gp|Z99109|BSUB0006_192 Bacillus subtilis complete genome (section 6 of 21): from 999501 to 1209940. NID: g2633260

10 . >gp:gp|Y09476|BSY09476_56 B.subtilis 54kb genomic DNA fragment. NID: g2145361.

gtgatactgatggaagaagcactaaaaagatagtatcttaggcgctcttgaaatggtaata
 gatcctgagttagggatagatatcgtaatttaggttttagtatataaagttgatgttgat
 gatgaagggtttatgtacagttgaaatgacattgacttcgatgggatgtccattaggacca
 caaattattgaacaagttaagagtggtttggctgagattcctgaaatttctgatacagaa
 15 gtgatgattgtatggagtccaccttgaataaagatatgatgtcacgatatgccaaaata
 gcttttaggcacgcgataa

Sequence 158

VILMEEALKDSILGALEMVIDPELGIDIVNLGLVYKVDVDEGLCTVEMTLTSMGCPLGP
 20 QIIEQVKSVLAEIPEISDTEVMIVWSPFWNKDMSRYAKIALGIG*

Sequence 159

Contig_0451_pos_16480_17565,

is similar to (with p-value 8.0e-75)

25 >gp:gp|U93874|BSU93874_12 Bacillus subtilis cysteine synthase (yrhA), cystathionine gamma-lyase (yrhB), YrhC (yrhC), YrhD (yrhD), formate dehydrogenase chain A (yrhE), YrhF (yrhF), formate dehydrogenase (yrhG), YrhH (yrhH), regulatory protein (yrhI), cytochrome P450 102 (yrhJ), YrhK (yrhK), hypothetical protein YrhL (yrhL), putative anti-SigV factor (yrhM), RNA polymerase sigma factor SigV (sigV) and YrhO (yrhO) genes, complete cds, and YrhP (yrhP) gene, partial cds. NID: g1934604. >gp:gp|Z99117|BSUB0014_194 Bacillus subtilis complete genome (section 14 of 21): from 2599451 to 2812870. NID: g2634966.

35 atgcctggattagatgggtttgcgagcaattgcagtcattggtattattattatcacttg
 aataaacaatgggttaacaggtggttttttaggcgtagatactttttttgtatttcaggt
 tatttgattacgagcttattacttaaaagagtatgaagatactggaacaataaatcttaaa
 aatttttggattcgtcgtattaaaaaggttattaccagcggtatttgcattaatagtagta
 40 gttggaattgcaactttattattgcacccgagcatattgtaagagttaaacatgatatg
 atagcagcaaatattttacgtatctaatgttggttatattgctaaagatgtcaattatttc
 gagcaaattttctttatgcctttaaagcacctatggtcactagccattgaagagcagttt
 taccttttttccagcagcagtactcttattattttatggcaatagtttaagaaaaagaaaaat
 gtcatactgatgttttgatcatatccctgggtttcattattaatgatggttgttatttct
 45 caacctcacttgaaccattctagagtatatatttggaaactgatacaagattgcagacactg
 ctttttaggtgtacttctagcatttatctggccaccttttaattaaatcccaatccacct
 aaaggattaaaaactgtgattaatagtcgggtatcataggacttacatttttaattcta
 ttattctttactgttagtgatgaaagtgattggatttataacggtggattttatcttatt
 tcaacaatgactttgctaattattgcaagtgttgttcatccaacgacaatttttagctaag
 50 ttattaggaaatcctttatttgtctacattggaaagcggtcatacagtttatacttatgg
 cattttcctgtaattagctttattcatagttattttattgatgggtcaattaccaacttat
 gtttatattatggatatcgtaattactgtattatttagccgaattatcatttagatatgtt
 gaaacgccattaagaaaggaaggtctaagggttttacagtgtgctcccttaaaaattat
 ttttag

55

Sequence 160

MPGLDGLRAIAVIGIIIIYHLNKQWLTGGFLGVDTFVVISGYLITSLLLKEYEDTGTINLK

NFWIRRIKRLLPVAFALIVVVGIATLLLHPEHIVRVKHMIAAIFYVSNWWYIAKDVNYF
 EQFSFMPKHLWLSLAIEEQFYLFPPAVLLLEMAIVKKKNVILMFWIIISLVSLMMVVIS
 QPHLNHSRVYFGTDTRLQTLGLVLLAFIWPPFKLNPNPPKGLKTVINSAGIIGLTFLIL
 LFFTVSDESOWIYNGGFYLISTMTLLIIASVVHPTTILAKLLGNPLFVYIGKRSYSLYLW
 5 HFPVISFIHSYFIDGQLPTYVYIMDIVITVLLAELSFYVETPLRKEGLKFTVCSLKNY
 F*

Sequence 161

Contig_0451_pos_17821_0,
 10 is similar to (with p-value 6.0e-46)
 >sp:sp|P49022|PIP_LACLA PHAGE INFECTION PROTEIN. >gp:gp|L146
 79|LACPIP_1 Lactococcus lactis pip and gerC2 genes, complete
 cds's, and rrg gene, 5' end of cds. NID: g308860.
 atgaaaaacgcactaaaaactttttatcacggatttataaaagagttgctaaaaacaccaggt
 15 gtatgggtcatcttagctggttttagcaattcttccttcattctatgcatggtttaacctc
 tgggctatgtgggactccgtatggtcatacaggacatatcaaagttgccgtagtgatgaa
 gaccaaggtgaaaaagttcgtggttaagaatattaatgtaggaaataaaatgggtcaaaact
 taaaaaagaatgatagttttgactggcaatttgtgagtagagaaaaagccgaccatgaa
 attaagatgggaaaatattatgcaggtatttatataccgaagaaattcacacatgaaatc
 20 actgggtactttaagaaaacatcctcaaaaggcggatatagattttaaaagtaaatcagaag
 attaatgctgtagcagctaagtttaaccgatacgggatcgctcgtttgtagttgataaagca
 aataaacaatttaacaaaaccgtagcaaccgctttactttctgaagctaataaagtcgga
 ctatcaattgaagataatgtacctacaatcaataaaattaagagtgctgtatatcaagct
 aataattcattgcctaaaattaatcaatttgcagacaagattattgaactaaataaacat
 25 caagacgatttgtagtgccttatgctaataatttagaagtttaggaaagtat

Sequence 162

MKNALKLFITDLKRVAKTPGVWVILAGLAILPSFYAWFNLWAMWDPYGHTGHIKVAVVNE
 DQGEKVRGKNINVGKMKVTKLKNDSFDWQFVSREKADHEIKMGKYYAGIYIPKKFTHEI
 30 TGTLRKHPQKADIDFKVNQKINAVAAKLTDTGSSSFVIDKANKQFNKTVATALLSEANKVG
 LSIEDNVPTINKIKSAVYQANNSLPKINQFADKIIELNKHQDDLDAYANQFRSLGKY

Sequence 163

Contig_0451_pos_16054_15662,
 35 is similar to (with p-value 2.0e-38)
 >gp:gp|Z79580|BS168NPRB_7 B.subtilis nprB gene. NID: g162092
 1. >gp:gp|Z99109|BSUB0006_192 Bacillus subtilis complete gen
 ome (section 6 of 21): from 999501 to 1209940. NID: g2633260
 . >gp:gp|Y09476|BSY09476_56 B.subtilis 54kb genomic DNA frag
 40 ment. NID: g2145361.
 atgagtagctatcctttatattatccgatgcctaaagctattttggcatatcgtgacatc
 atatctttattccaaggtggactccatacaatcatcacttctgtatcagaaatttcagga
 atctcagccaaaacactcttaacttgttcaataatttgggtcctaattggacatcccatc
 gaagtcaatgtcatttcaactgtacataaaccttcatcatcaacatcaactttatatact
 45 aaacctaaattaacgatattctatccctaactcaggatctattaccatttcaagagcgcct
 aagatactatcttttagtgcttcttccatcagtatcacctctttaaaattttctttacac
 caatatatcaaatatccgacaaaacgccaataa

Sequence 164

50 MSSYPLYYPMPKAILAYRDIISLFQGGHLTIITSVSEISGISAKTLLTCSIICGPNGHPI
 EVNVISTVHKPSSSTSTLYTKPKLTISIPNSGSITISRAPKILSFSASSISITSLKFS
 LYH
 QYIKYPTKRQ*

55

Sequence 165

Contig_0451_pos_15613_14792,
 is similar to (with p-value 1.0e-74)
 >gp:gp|Z79580|BS168NPRB_5 B.subtilis nprB gene. NID: g162092

1. >gp:gp|299109|BSUB0006_190 *Bacillus subtilis* complete genome (section 6 of 21): from 999501 to 1209940. NID: g2633260
 . >gp:gp|Y09476|BSY09476_54 *B. subtilis* 54kb genomic DNA fragment. NID: g2145361.
- 5 atgcaaccttattttaatttgtctagatctagatggtacattattaaatgacaataaagaa
 atctcaccttacactaaacaagtatttaaccgaattacaacaatgtggacactacgttatg
 attgctactggaagaccctatcgcgcaagccagatgtattatcatgaactaaatagagc
 acacctgttggttaactttaatggagcatttgtacatcatccaaaagcaaacgattttaaa
 gtgatacatgaagtacttgatgtggaaatttctaaaaatattattacagcacttcaacaa
 10 tctcatattacaaatatcattgctgaagtaaaagactatgtctttataaatagttatgat
 tcaagactttacgaagggttttcaatgggaaatcctaaaaattcaaacaggttaatttactt
 gaaaatcttaaatgaagcacctacgtcattacttgttgaagcagaagaagaaaatttcct
 gaaattaaagatatgttaacacatttttatgcagaaaattgaacatcgtcgttggggc
 gcaccgtttccagtaataagaaattgtgaagcgtgggattaacaaagcacgtggaatcaag
 15 catgttcaaaactattttaacatcgccgacgatcatatcattgcgtttggtgatgaggac
 aatgatataagaaatgataaagtttgcgacccatggcattgcaatggccaatggcttgaaa
 gatttaaaggaaatagcaaatgagactacgtatagtaaatgaagacggaatagggtcgt
 tattttaaatgacttttttaatttgaaaatacgttattattaa
- 20 Sequence 166
 MQPYLICLDLDTLLNDNKEISPYTKQVLTELQCCGHYVMIATGRPYRASQMYHELMNS
 TPVVNFNGAFVHHPKANDFKVIHEVLDEISKNIITALQQSHITNIIAEVKDYVFINSYD
 SRLYEGFSMGNPKIQTGNLLENLNEAPTSLIVEAEEENIPEIKDMLTFYAENIEHRRWG
 APFPVIEIVKRGINKARGIKHVQNYLNIADDDHIIAFGDEDNDIEMIKFATHGIAMANGLK
 25 DLKEIANETTYSNNEGDIGRYLNDFFNLKIRYY*

- Sequence 167
 Contig_0451_pos_14714_13398,
 is similar to (with p-value 0.0e+00)
- 30 >gp:gp|AF041467|AF041467_1 *Staphylococcus aureus* coenzyme A
 disulfide reductase gene, complete cds. NID: g2792489.
 atgaataaaattataatagtcggtgcaagtgtgctggtggtgcgacttgtgcaagtcaaatt
 cgaagattagataaagagagtgaaatcattgttttgaaaaagatagagacatgagcttt
 gctaattgtgcattaccttattatattggcaacggttatcgaggaccgtcgttaaagtttta
 35 gcatacacgcccattcaattttatgacaaaaagcaaactactgtaaaaacataccatgaa
 gttatacaaatcaatgatgagagacaaacagttactgtcttaaatcatcaaactaatcaa
 acttttgaaagaaagtacgatacattgattttaagtccctggcgcatctgcaaatcgatta
 aacactcatagtgatatctcatttactgtgcgaaatctcgaagatactgaaacaattgat
 acctttattacgaataccaaagcacaacgtgcacttgttgttggcgcggttacatctct
 40 ttgaagtccttgaaaatttacatcatagaggtttggatgtcacatggattcatcgctct
 acaaatattaataaactgatggatcaagatatgaatcaacccatcatcgacgaaatagaa
 aagagaaatatcacttatagatttaacgaagaaattagtcacgtaaatggacatgaagtt
 acattcacatctggttaaagttgaaaactttgatcttattatcgaagggttaggtactcat
 ccaaatccacaatttattaatatcatctaacgtcatactgaatgataaagggttatatccca
 45 gtaaatcataatttccaaacaaatataccaaatatttatgcattaggtgatgttattact
 tcacattatcgctcatgtgaatttaccggcacaggttccacttgcttggggagcacaccgt
 ggtgcaagtattatagctgaacaactttctggaaattcgtctattcactttaagggttat
 ctaggaaataatatagtgaattttttgactatacattagcaagtgttggcatcaaacca
 aatgaacttaaaaaatttcgattatgatatggttgaagtttaagcaaggagctcatgcagga
 50 tattaccaggaatttcaccactacatttacgtgtttattttgaaaaagactcgagaaaa
 ctatacgcgcagcagcaggttggttaacaagggtgccgataaaagaatagacgtattatca
 atggcaatgatgaataatgctactgtggatgatttaacagaatttgaaagtagcatatgca
 cctccttatagtcacaaaaagatttaatttaatttaattgggtataaagcgcaataa
- 55 Sequence 168
 MNKIIIVGAVAGGATCASQIRRLDKESEIIVFEKDRDMSFANCALPYYIGNVIEDRRKVL
 AYPNQFYDKKQITVKTYHEVIQINDERQTVTVLNHQTNQTFEESYDTLILSPGASANRL
 NTHSDISFTVRNLEDTEITIDTFITNTKAQRALVVGAGYISLEVLENLHHRGLDVTWIHRS
 TNINKLMDQDMNQPIIDEIEKRNTYRFNEEISHVNGHEVFTTSGKVENFDLIEGVGTH

PNSQFIKSSNVILNDKGYIPVNHNFQTNIPNIYALGDVITSHYRHVNLPAQVPLAWGAHR
 GASIIAEQLSGNSSIHFKGYLGNNIVKFFDYTLASVGIKPNELKNFDYDMVEVKQGAHAG
 YYPGNSPLHLRVYFEKDSRKLIRAAAVGKQGADKRIDVLSMAMMNATVDDLTEFEVAYA
 PPYSHPKDLINLIGYKAQ*

5

Sequence 169

Contig_0452_pos_547_1323,

is similar to (with p-value 8.0e-40)

>sp:sp|Q57629|Y165_METJA HYPOTHETICAL PROTEIN MJ0165. >gp:gp
 10 |U67473|U67473_9 Methanococcus jannaschii section 15 of 150
 of the complete genome. NID: g2826256.

atgagccatagttataattctatagaagaggtgctcaaagctgtaaaatcaaactcaacta
 tctattaatgatgctaaagcccaactcagtcattatgacgaattgggctttgctaaaatt
 gacttacatagagcacagcgtcaaggatttcccgaagttatctttgggcaaggaaaaaca
 15 aaagaacaaatcactaaaatcatctctagtttgatatttcataatgaagttattctagtg
 acacgtgttgatgaaatgaaagcaaaatacattttacaacattatccaaacttggaatat
 catcaaactgcacagtttaattagcactccactaaaagatataccacaatctaaatactat
 gtttctgtactttgtgctggaacttctgatttacctattgcagaagaagctgcattaacc
 gctgaaatcatgggagtaagtgtaaaacgattttatgatgtcggggtttcaggtattcat
 20 cgcttattatccaacattcatgatatacgcagaggggaaagtttctatcgttatagctgga
 atggaaggcgcttttagcaagtgttggaggagattagtcaccaccctgtatatgcagta
 ccaacgagtgtaggttatggagcaaaactggaatggggttaccaccctattatcaatgata
 aatagttgcgcacccggaaccagcgtattaaatatcaataatggatttgggtggcggttac
 aacgctgcacagattattcatatgctagaaaataaagagagtgaggtatctttatga

25

Sequence 170

MSHSYNSIEEVLKAVKSNQLSINDAKAQLSHYDELGFAKIDLHRAQRQGFPEVIFGQGKT
 KEQITKIISLIFHNEVILVTRVDEMKAKEYILQHPNLEYHQTAQLISTPLKDIPOQSKYY
 VSVLCAGTSDLPIAEAAALTAEIMGVSVKRFYDVGVSIGIHRLLSNIHDIRRGKVSIVIAG
 30 MEGALASVVGGLVNHVPVYAVPTSVGYGANLNGVTLLSMINSCAPGTSVLNINNGFGGGY
 NAAQIIHMLENKESEVSL*

Sequence 171

Contig_0452_pos_1368_2507,

35 putative peptide of unknown function

atgctactttctgcttttagttgatttaggagcaaaccctgaagacattgaatcagaacta
 aaaaaattaccttttagatcaatttaagctacattttcaaaaaagagtaaaacaaggtatt
 catgcaatgacattaaacattgatgttaagaagcaaatcatcatcgctcacgttaatgat
 atatttaaaatgatagatgacagtagacttccggaaagggttaaataatcgagtaagaaa
 40 atttttgaaatcattgggtcaagcagaagctaaaattcatggcatgtcgtttgaagaagtt
 cactttcatgaagtgggggcaatggactctattatagatatatttgggtgggtgtattgca
 ctagaacaactagggattaacacattatactgttcagctattccaacaggtcatggtaaa
 atcaatattgctcatggcatttatccaatccctgcaccagctactgctgaaattctttaa
 ggtataccaatcgacattttgatgttcaaagtgaactcacaacccctactgggtgctgca
 45 tttgctaagggaacttgtttcatcggttgggccatttccctcagcaacaatacaacatata
 ggctattggcgccggcagtaaggattttgatttccctaataatattaagggttattcaattt
 gaatctgaattcgagcaacaagatagcgtccaagtaataagagtgcaaatagatgatag
 acacctgaagcattaggttattttatgaataatgcgttagagcaagggtgcttttagatgct
 tactatacgcttatattttatgaaaaaagtcgcccagcagcaggttaacgttaatatgt
 50 aaattacatgataagacatatcttgaacaacttatcttacaagaacaagttcttttaggc
 gtcagaagtacttctgttaatagaaagaccttgaaccgcgattcaaaattctttctaca
 caacacggcactgtttccattaaatttggcctacaaaatggaaaaattatgaaaaatgaaa
 cccgagtatgaagatttgaagaaaatagctaaaactacaaaacaaccgtttcaagtaatt
 cataacgaggtattacaacaactctatcaaacatatcatatagggaatatatttcaataa

55

Sequence 172

MLLSALVDLGNPEDIIESELKKPLDQFKLHFQKRVKQGIHAMTLNIDVKEANHHRRHVND
 IFKMIDDSTLPERVKYRSKKIFEIIGQAEAKIHGMSFEEVHFHEVGAMDSIIDIIIGGCIA

LEQLGINTLYCSAIP TGHGKINIAHGIYPI PAPATAEILKGIPIAHFDVQSELTTPTGAA
 FAKGLVSSFGFPFSATI QHIGYGAGSKDFDFPNILRVIQFESEFEQQDSVQVIECQIDDM
 TPEALGYFMNNALEQ GALDAYYTPIFMKKS RPSTQLTLICKLHDKTYFEQLILQETSSLG
 VRSTSVNRKTLNRAFKILSTQHGT VSIKFLQNGKIMMKPEYEDLKKIAKTTKQPFQVI
 5 HNEVLQQLYQTYHIGNILQ*

Sequence 173

Contig_0452_pos_3958_3161,

is similar to (with p-value 1.0e-85)

10 >sp:sp|P39651|YWF0_BACSU HYPOTHETICAL 51.0 KD PROTEIN IN PTA
 3'REGION. >gp:gp|Z99123|BSUB0020_56 Bacillus subtilis compl
 ete genome (section 20 of 21): from 3798401 to 4010550. NID:
 g2636240. >gp:gp|Z80355|BSUWFO_1 B.subtilis ywfO, ywgA and
 ywgB genes. NID: g1561566.
 15 atgatttcctcacaaattgatgctgatcgaaatggactatttacaaagagatgcatat
 acaggcgtaacgtatggctcatttgatggagcgtattttaaggtgatgagaccatct
 aaagaagaagtgttaataaagatagtggtatgcatgctgctcgaaaattttattatgagt
 cgttatcaaattgtattggcaaatatatttcatccagtaagccgtgggtgggaagt
 ttaaacaattgtttaaacgagctaagcagctttataatgaaggatgatgaatttaaatg
 20 tatccaaaagactttataccattcttgaaggaacaatgacgattgaacaatatgtagaa
 ctgatgaagcagttgtattgtattacttgaagaaatggattcatgaaaatgatacaata
 ttaagtgatattatcaagacggtttatcaatcgagatttatttaaatatattcctttcgac
 ggttcaattattaccatttcggaattgcaagaattatttgaagcgggtggtattaatcct
 gattattactttgtaagtgaagcattttcagatttaccttatgattatgatcgccaggc
 25 tcaaatcgcaaacccgattcatttataaaaagtaatggtggaattacagaaataagtaat
 caatcattggtgatataatgattacagggattaatagagaagaccataaattatattat
 cctaaagagatgatttttaaaaattaaagattatcaaattaaaggttctattattaactta
 cttaataatgaattaaataa

30 Sequence 174

MISSQIDADRMDYLQRDAYFTGV TYGSFDMERILRLMRPSKEEVLIKDSGMHAVENFIMS
 RYQMYWQIYFHPVSRGGEVLLNNCLKRAKQLYNEG YEFKMYPKDFIPFEGTMTIEQYVE
 LDEAVVLYLKKWHENDTILSDLRRFINRDLFKYIPFDGSIITISELQELFEAGGINP
 DYYFVSEAFSDLPYD YDRPGSNRKP IHLKSNNGGITEISNQSLVINSITGINREDHKLYY
 35 PKEMILKIKDYQIKGSIINLLNELN*

Sequence 175

Contig_0453_pos_1385_2191,

is similar to (with p-value 5.0e-38)

40 >gp:gp|D86240|D86240_5 Staphylococcus aureus gene for unkown
 function and dlt operon dltA, dltB, dltC and dltD genes, com
 plete cds. NID: g1405333.
 atgtcacaaagtcaaatcaatcagatgtttaatcaaaaagatatgccagctaatttgaag
 aaacggtatgcacaaagattgttacagtttccgcatgcacacaataagtcataccttaga
 45 gaacaagcaaaacatcctaataatgatgtctctggaaactacatttcttcatttaagaaaat
 caattaactaagattgaagctattaaatcattattctcattcactaagccacctctagca
 gaagtaaaacctgcaacaagagaagatgcttcatgggatgagatgaaacataaagctgcc
 gatataggcaaaagcaaaataactcaatctaataaatatgatataagagatccatattggaaa
 ttgataaaacaaaacaaagcgtaaaaatcaaaagggattatgagttcaacattaactcacc
 50 gagttccaagattttaaattatttagtgcaaacgctacatgctgctggagctgatgtacaa
 tatgtttgtataccttcaaatggaagatgggtatgatcatataggtatcaaaaaagataga
 cgtgaagctgtatataaaaagattcactcaactgtagttgataatgggtgggaaat
 gatttgacaaataaggactatgaaaagtacgtaattagtgatgctgttcatttgatgg
 aaaggttgggttacgtcgaccagcaaatgcaagacatatggatgggtcatgcgcctaaa
 55 aatcatgaagtcgattatttctaaaaataaaccaccgcacaaacatcacacgatcgtcaa
 gatgatcaacatcaaggcaacaaataa

Sequence 176

MSQSQINQMFNQKMPANLKKRYAQRLQFP HAHNKS YLREQAKHPNDVSGNYISSFKEN

QLTKIEAIKSLFSFTKPPLAEVKPATREDASWDEMKHKAADIGKANTQSNKYDIRDPYWK
 LIKQNKRIKRDYEFNINSPEFQDLKLLVQTLHAAGADVQVCIPSNRWDHIGIKKDR
 REAVYKKIHSTVVDNGGKIYDLTNKDYEKYVISDAVHIGWKGWVYVDQQIARHMDGHAPK
 NHEVDYSKNKPPHKHHNDRQDDQHQNK*

5

Sequence 177

Contig_0453_pos_2831_3151,

putative peptide of unknown function

atgactaaaattagtggtgtcgtatatggagcagaagtcgtttgtgcgagttgtgtaaat
 10 gcacctacatctatagatacttatcaatggcttcaagcattacttttaagaaagtttcct
 caacatcattttgaatttacatataattgacatacgaaatgatactgaaaatttaactgat
 catgatatgcaatttatagaaagaattaatgaagatgaattgttttaccattagttacg
 atgaatgatgaatatgtagcagatggttacatacaatataaacaataaaccggttttatt
 aatcatattttactatgtaa

15

Sequence 178

MTKISVVVYGAEVVCASCVNAPTSIDTYQWLQALLLRKFPQHHFEFTYIDIRNDTENLTD
 HDMQFIERINEDELFYPLVTMNDYVADGYIQYKQITRFIKSYFTM*

20 Sequence 179

Contig_0453_pos_0_376,

is similar to (with p-value 1.0e-43)

>gp:gp|Y09570|SAFEMD_1 S.aureus femD gene. NID: g1684748. >g

p:gp|Y15477|SAARGFEMD_4 Staphylococcus aureus argI, glmM gen

25

es and ORF1 and ORF2. NID: g3892891.

atgccatctattccagaaatctttaatatTTTTGGCTTTAAACGGTTGTTAAAGTTAGG
 cggtgtcttaattgattattaacgggtgttcagttggcttccgttattaaaatgagtggt
 aaaactctaagcgagttagcttctcaaatgaaaaagtaccacaatctttaattaatgtg
 agagtgactgacaaatatcgtgttgaaagagaatattcatgttcaagagataatgacgaaa
 30 gttgaaacagagatgaatggtgaaggaagaattcttgttcgtccttctggaactgaacct
 ttagtacgtgtaattggttgaggctgcaactgacgcggatgctgaaagatatgctcaaagt
 atcgtctgacCGCGACA

Sequence 180

35 MPSIPEIFNIFGKRLFKVRRCLNGLLTGVQLASVIKMSGKTLSELASQMKKYPQSLINV
 RVTDKYRVEENIHVQEIMTKVETEMNGEGRILVRPSGTEPLVRVMVEAATDADAERYAQS
 IADRD

Sequence 181

40 Contig_0454_pos_441_1559,

is similar to (with p-value 0.0e+00)

>gp:gp|Y14370|SAY14370_3 Staphylococcus aureus RF3, murE, yp
 fP genes. NID: g3256221.

45

atgcaagtcacgcaaagtattgtcaaccaattgaatgagatgaatctcaatcatttatca
 gtcatccaacatgatattgtttatggaagctcatccaattatgacttctatatgtaagaaa
 tggatatcaatagctttaaatatTTTAGAAATACATATAAACGATTTACTATAGTCGC
 cctaattgagctcgataaatgtttttataaatattatggattaaataaaactcatcaactta
 ctatttaaagaaaagcctgatctcatattattaacatttccaacacctgtgatgtcagtg
 ttgaccgaacaatttaataataaatatccctattgacgacagttatgacagattatcgcatg
 50 cataaaaattggattacaccatattcacaaagatattatgtagcaacaaaagatactaaa
 gatgatttcattgaagctggtgttcctgcttcatatattaaagtgcgggcattoctatt
 gctgataaaattgaaagaatctattgataaagaagaatggttatcgcaacaacatttagac
 ccttcaaaacctactatattaatgtcagcaggtgcatttgggtgtttcaaaaggctttgac
 tatatgattaataatatTTTAGAAAAAGTCCAAATTCGCAAGTGGTCATGATTGTGGA
 55 cgtagtaaggaaacttaaacggttcattaaaagctaagttcaaagataatccaagtgtata
 atattagatatatacaaatcacatgaatgagtggtatggcatcaagccaactaatgattaca
 aaacctggtggtatcacaaatttccgaaggacttagtcgttgatttcttatgatttttta
 aacctgcacccggtcaagaacttgaaaatgcatattactttgaaagtaaaggatttgga
 aaaatagcagatactccaatgaggcaattgatattgtttctgacttaacaaataacgaa

gagacttttaaagggttatgtcatctaaaatgctagaatcaaaggtaggatattctactaga
aagatttgtaaagatttattagatttaaataggctactcatctcaaccggatgaaatctat
ggaaagggttcctttgtatgcaagattcttcgtcaagtaa

5 Sequence 182

MQVTQSIQVNLNEMNLNHLNLSVIQHDLFMEAHPIMTSICKKKWYINSFKYFRNTYKRFYYSR
PNELDKCFYKYYGLNKLINLLIKEKPDILLTFPTPVMSVLTEQFNINIPATVMTDYRM
HKNWITPYSQRYVATKDKDDFIEAGVPASYIKVTGIPIADKFEESIDKEEWSQQLD
PSKPTILMSAGAFVSKGFDYMINNILEKSPNSQVVMICGRSKELKRSKAKFKDNPSVI
10 ILGYTNHMNEWMASQLMITKPGGITISEGLSRCIPMIFLNPAPGQELNAYYFESKGF
KIADTPNEAIDIVSDLTNNEETLKVMSSKMLESKVGYSYTRKICKDLLDLIGHSSQPDEIY
GKVPLYARFFVK*

Sequence 183

15 Contig_0454_pos_1585_2730,

putative peptide of unknown function

atgatgttggaattctgtttctgatggaatttgcaagaggtatgtacataactaagttat
ataaactttttacctacagtgacctctatcgcaatagcaatcacatcatttgctttttcc
attcactttatcgagatgctgcaacaaattttgtcatcggtttttacttaaaaaattt
20 gggtcaaaattatgacttactacatctggattcttacttgcttttataagcttggttttagtg
atatgggtcccgcatcaccattcataatttttccagtgtctattatgttaggaattgct
gtgagtcgatttggttatcatgttatctagtgtagatgaaagaaatcgcggaacaa
atgggttatgtctacttttcatggttgctaggtttattgggtgggtatggttatcatgaac
ttgcttattaaattccatcctactcgttttgcatttttaaaggccttggttgcttatt
25 gcttgggtactatactattttgttaatatcaacttaacaaattacaataactaaacctgtg
aaagcacaattaaagcaaattgtagatgttacacaacgtcatcttatttctatttccgggt
atcttggttacaaggagcagctatagcagcacttgtaacctattcttccaaaatatgcaacg
caagttgtgaaagtatcaaccgttgaaatatacagtagcaatcattattgggtggcataggc
tgtgctttctctatgttatttttatcaaaaatcatcgacaataatagcaaagggtttatg
30 tatggagttatttttagtggtttatactatatacaattcttatattcggttatctaca
attacaaatatatatagtttgggccataggactttttattgggttaattgacgggtatc
ctcttaccggcttggaataacctttatggctgggcatattaatcctaacgaacaggaagaa
acatggggcgtgttcaacagtggtcaaggcttcgggtcaatgataggccactagtcgga
ggtctaattactcaatttactaataatttaataataaccttttacttttcagcgatgatt
35 tttcttgcacttgcatgtttttacggatattactttattaaaacaaacagaagggttaaa
ccttaa

Sequence 184

40 MMLVILFLMEFARGMYILSYINFLPTVTSIAIAITSFAFSIHFIADAATNFVIGFLKKF
GSKLVLTSGFLAFISLFLVIWFPASPIIIFSAIMLGIAVSPWVIMLSSVDERNRGKQ
MGYVYFSQLGLLVGMVIMNLLIKFHPTRFAFLMALVVLIAWVLYFVNINLTNYNTKPV
KAQLKQIVDVTQRHLILFPGILLQGAAIAALVPILPKYATQVVKVSTVEYTVAIIGGIG
CAFSMLFLSKIIDNNSKGFMYGVIFSGFILYILIFGLSTITNIYIVWAIGLFIGLMYGI
LLPAWNTFMAGHINPNEQEETWGVFNSVQGFSGMIGPLVGGILITQFTNNLNNTFYFSAMI
45 FLALAVFYGYFIKTNRRVKP*

Sequence 185

Contig_0454_pos_4142_0,

is similar to (with p-value 3.0e-66)

50 >gp:gpIU57060ISAU57060_1 Staphylococcus aureus scdA gene, co
mplete cds. NID: g1575060.

atgaaattcttccctacgtttttcaacatcttcgtaattggaattttgttggtggcggaacaa
gagagtatcgcttcagctgtcaatcataaaccacaaatttgacttaaaattccttattaaat
aagttgaatcatattgataatacagaaggtaacagtagcatttaattcctaaatttttaaa
55 gttgaatctcttatataatatacaatcagcttatcacgaaacgcttaagaagaattt
aagaatcttacaccttacatgactaaattggcaaaagtacatggtcctagtacccatac
ttattaaaattacaagacttatatcgcgagtttcgtgatagtatgttggtatcatatacgt
aaagaagatgaggaagattttcctaaactcattcaatatagtcaggacaagatgtacaa
aacattaaaatcatattagaagatttaattaatgaccacgaagatactgggcaattatta

aatgttatgaatcaactaacctctgattatcaaaccacagaagaagcatgtggaacatgg
aagcttggtttaccaaagattacaaaatatcgacgtcaaacacaccaacatgtacat

Sequence 186

5 MKFFPTFFNIFVMEFCCGGQESIASAVNHKPNIDLNSLLNKLNHIDNTEGNSTINPKFLN
VESLIQYIQSAYHETLKEEFKNLTPYMTKLAKVHGSPHPYLLKLQDLYREFRDSMLDHIR
KEDEEDFPKLIQYSQGQDVQNIKIILEDLINDHEDTGQLLNVMNQLTSDYQTPEEACGTW
KLVIYQRLQNIERQTHQHVH

10 Sequence 187

Contig_0454_pos_4111_3662,

putative peptide of unknown function

atgggaaaagaaattcttcttatatcgatgcgacatttccaacttataaagtaggtaat
acaaggttacttatttgagatagtttagcaggaagtatcgctttaatgactgcaatgact
15 taccacaactattttttagtgcaggttgcttattgagcccaatgtataatgaaaatattaag
aaaaaaattgatacatgtatgaataaaggtaattgacgatatggcatgccattgggtta
gaagaagcagattttttttaccaactaatggtaaaagagctaactttttaacacctaac
cgtgaattaaatcaactgattaaagaagataatattgaatatttctataaagaatttaac
ggtggacatcattggaaatcatggaaaccattgctaggagatattctcttacaattttta
20 ggtgatccaataaatggaaaatatgtttta

Sequence 188

MGKEILPYIDATFPTYKVGNTRELLIGDSLALMTAMTYPTIFSRVALLSPMYNENIK
KKIDTCMNKGQLTIWHAIGLEEADFILPTNGKRANFLTPNRELNLQLIKEDNIEYFYKEFN
25 GGHWWKSWKPLLGDILLQFLGDPINGKYV*

Sequence 189

Contig_0454_pos_3509_3000,

putative peptide of unknown function

atgatttttaggattagcattgggtccgtcaaagtcatttcaagatgaggtgaatgcttat
cgcaagcgatattgacaatcattatgctcaataatgcctcatatcacgattaaacctcaa
tttgaaatcgatgatcatgatttttaatttaataaaaaatgaagtgaataatcgaatttct
agtattaaaccagtagaagtagatgctacaaaggcatcctaatttcgctccaatcagtaat
gttatatacttcaaaagttgctaaaacagagtcattagatcaattatttaataatcaattta
35 acagaagattttttacggtacagctgaacatccttttgaccacattttacaattgccc
ggtctaacaagtcagaattttgaagatatatatggtcaagtaaaattagcaggggtagac
catagagaaataattgaagaactatcgttacttcaatatagtgagaagaggacaaatgg
actattattgaaacttttacattaggataa

40 Sequence 190

MILGLALVPSKSFQDEVNAYRKRYDNHYAQIMPHITIKPQFEIDHDFNLKNEVKNRIS
SIKPEVHATKASNFAPISNVIYFKVAKTESLDQLFNQFNTEDFYGTAEHPFVPHFTIAQ
GLTSQEFEDIYQVKLAGVDHREIIIEELSLQYSEEDKWTIIETFTLG*

45 Sequence 191

Contig_0455_pos_5713_5009,

is similar to (with p-value 2.0e-38)

>sp:sp|P17166|TRPA_LACCA TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC
4.2.1.20). >pir:pir|S42347|JS0344 tryptophan synthase (EC 4
50 .2.1.20) alpha chain - Lactobacillus casei >gp:gp|D00496|LBA
TRP_6 Lactobacillus casei DNA, trp operon (trpD, trpC, trpF,
trpB, trpA), complete cds. NID: g216754.

atgggtgatttaaaattttattcatcatttataaaacatttaactgagaatggagcagacatt
gttgaaattgggtgtgccattttctgacatcctgttgagatggacctataatcatgaaagca
55 gggcgcaacgctattgacgaggttcaaacatttaaatcatttttgatgaattataaaaa
aataaaaaatactattttcatcctaagtatgtattaatgacttattataatatttctaagtgt
tatggagaagaattatttttgataagtgatgaagctggtgtttatggtttaattatt
ccagatttaccttcaaaccttcaaaaaagtttaaaaaagatttttatcatcattctggt
aaaataatatcgtttaattgccatgaccgcaagtgatgctaggattatgcaaattgcaaag

5 aactcagaaggattttatttacacggtaacaatgaatgccacaacaggttaacagtgaggag
 ttccatccagatttaagagaaaaattgaatatataaaaaaagtttcaaaaattcctgtg
 gttgctggatttggtatcaaaaatcctgaacatggttaagatatagcgtccggtgcagat
 ggtattgtaattggtagtgaattgtaaaacgtattgaaatagattcaagaaaagaattt
 atcacttatatcaaatcaataagaactacgttgaaattctttataa

Sequence 192

MGDLNFIHHLKTLTENGADIVEIGVPFSDPVADGPIIMKAGRNAIDEGSNIKFI FDELIK
 NKNTISSKYVLMYYNILSAYGEELFLDKCDEAGVYGLIIPDLPYELTKKFKKDFYHHSV
 10 KIISLIAMTASDARIMQIAKNSEGFYITVTMNATTGNSGEFHPDLKRKIEYIKKVSkipV
 VAGFGIKNPEHVKDiasVADGIVIGSEIVKRIEIDSRKEFITYIKSIRTTLNSL*

Sequence 193

Contig_0455_pos_3811_2633,
 15 is similar to (with p-value 0.0e+00)
 >gp:gp|U23713|SEU23713_1 Staphylococcus epidermidis factor e
 ssential for methicillin resistance FEMA (femA) gene, comple
 te cds. NID: gl815617.
 atggaaggtaattacgaattaaaggttgctgaaggtaccgagtcacatttagttggaatt
 20 aaaaataatgataacgaagtgtgagctgtttatttaacagctgttcctgtaattgaaa
 atatttaaatatttttattccaatcgcggtccagtaataagattataataataagagctt
 gtacattttttctttaatgaattgagtaaatatgtaaaaaatataattgtttatattta
 agagttgaccataccttccatatacaatatttaaatcatgagggagaaaataactggaaat
 gcaggtcatgattggatttttgatgaattagagagtttaggatataaacacgaaggattc
 25 cacaaggatttgatcctgtattacaaatccgatatacattctgttctaaatttagcaaac
 aaaagtgtcaatgatgttttaaaaaacatggatggtttaagaaagcgtaataactaaaaaa
 gtttaagaaaaatggagttaaagtcgcgtttttatctgaagaagagttacctatatttagg
 tcattttatggaggtacacctctgaactaaagattttgcagatagagaagatagttttat
 tacaacagattcaaacattataaagaccgtgttttagtaccactagcctatattaacttt
 30 gatgagtatatagaggaaactaaataatgaagaaatgtgcttaataaagattataataaa
 gctttaaaagacattgagaaacgtccagagaataaaaaagcacataacaaaaggaaaat
 ttagaacaacaactcgatgcaaatcagcaaaaaatgaagctaaaaacttaaaacaa
 gaacatggcaatgaattacccatctctgctggcttctttataattaatccgtttgaagta
 gtttactacgctggtggaacttcaaatcggttatcgccattttgcaggagctatgcggtt
 35 caatggaagatgattaaactatgcaattgaacatggtattaatcggtataatttctatggt
 attagtggtgacttttagtgaagatgctgaagatgctggcgtagtttaagtttaaaaagggc
 tatgatgccgatgttatagaatacgttggtgactttattaaacctattaataaaccaatg
 tataacattttatagaacacttaaaaaactaaagaaatag

40

Sequence 194

MEGNYELKVAEGTESHLVGIKNNDNEVIAACLLTAVPVMKIFKYFYSNRGPVIDYNNKEL
 VHFFFNELSKYVKKYNCLYLVRDPYLPYQYLNHEGEITGNAGHDWIFDELESLGKHEGF
 45 HKGFDPVLQIRYHSLVNLANKSANDVLKNMDGLRKRNTKKVKKNGVKVRFLSEELPIFR
 SFMEDTSETKDFADREDSFYNNRKFHYKDRVLVPLAYINFDEYIEELNNERNVLNKDYNK
 ALKDIEKRPNKKAHNKKENLEQQLDANQQKINEAKNLQEHGNELPISAGFFIINPFV
 VYYAGGTSNRYRHFAGSYAVQWKMINYAIEHGINRYNFYGISGDFSEDAEDAGVVKFKKG
 YDADVIEYVGDFIKPINKPMYNIYRTLKKLKK*

50

Sequence 195

Contig_0455_pos_2607_1354,
 is similar to (with p-value 0.0e+00)
 >gp:gp|U23714|SEU23714_1 Staphylococcus epidermidis factor e
 55 ssential for methicillin resistance FEMB (femB) gene, comple
 te cds. NID: gl815619.
 atgaaatttacagagtttaacagtttaagaatttgaaaactttgtacaaaatccatcatta
 gaaagtcattattttcaagtgaaggaaaatattgctacacgtgaatcagatgggtttcaa
 gtagtgttattaggtgtaaaagacgacgacaatagagtgatagcagctagcctgttttct

5 aaaatccctacaatgggcagttatgtgtattattccaatagaggccctgtaatggactat
 tcagatttaggttttagtggatttttattttaaagagcttgataaatatttacatcaacat
 caatgcttatatgtaaaattagatccttactggttgatcaagtttatgataaagatatt
 aatcctttaacagaaaaaaatgatgcttttagtaaattctatttaaatacacatggttatgat
 catcacggatttacaacccaatatgattcttccagccaagttagatggatgggggtatta
 10 gatttagaaggcaaaaccctgcattctctaaggaaagagtttgatagtcaaaagaaacga
 aatattaataaagcgataaaactacgggtgtgaaagttagatttcttagtaaggatgaattt
 gatttattcttagacttataccgagagactgaagctagaactggatttgcttctaaaact
 gacgattatttctataactttatagagcattatggcgataaaagtattagttccttttagct
 15 tacatagattttaaataatataatacaacatttgcaagaatcactaaatgataaagaaaaat
 cgacgtgatgatgatgaggcgaagaaaaataaaacagataaaacagttaaagaaaaatagct
 gagttagataaaacaaattgatcacgataaaaaagaattgcttcaagctagtgaattacgt
 caaacagatggcgaaattttaaattttagcttcaggagtatactttgctaattgcatatgaa
 gtgaactatttctctggagggtcttcagaaaaatataatcaatatatgggaccatattgca
 20 atgcattggcacatgattaattattgttttgataacggttatgatagatataatttctat
 ggcttatcaggtgattttactgaaaacagtgaaactatggtgtttatcgctttaagaga
 ggttttaattgtaggattgaggaattaatcggtgatttctataaaaccaatcaataaagtg
 aaatattggttattcaatacatttagatcgcatacgtaataaattgaaaaagtaa

20 Sequence 196

MKFTELTVKEFENFVQNPSLESHYFQVKENIATRESDFQVVLGVKDDDNRVIAASLFS
 KIPTMGSYVYYSNRGPVMDYSDLGLVDFYKELDKYLHQHQCLYVKLDPYWLQVYDKDI
 NPLTEKNDAVLNLFKSHGYDHHGFTTQYDSSSQVRWGMVLDLEGKTPASLRKEFDSQRKR
 NINKAINYGKVRFLSKDEFDLFLDLYRETEARTGFASKTDDYFYNFIEHYGDKVLVPLA
 25 YIDLNEYIQHLQESLNDKENRRDDMMAKENKTDKQLKKIAELDKQIDHDKKELLOASELR
 QTDGEILNLASGVYFANAYEVNYFSGGSSEKYNQYMGPHYAMHWHMINYCFDNGYDRYNYF
 GLSGDFTENSEDYGVIYRFKRGFNVRIEELIGDFYKPKINKVKYWLFTLDRIRNKLKK*

30 Sequence 197

30 Contig_0456_pos_6598_8601,
 is similar to (with p-value 1.0e-81)
 >gp:gp|AE001272|AE001272_20 Lactococcus lactis DPC3147 plasm
 id pMRC01, complete plasmid sequence. NID: g3582195.
 35 gtgacaaatgcaacgcctgaacaatataacccttcataataaagaatggaatttagaagac
 ttacctatcatttctaagaaaatgaaaacagtagtgatttagtaaaacaaatagacaattt
 aaaattgtaaaatctttaatttttagataaaaatgtaaagaaattattatagcaacagat
 gctggacgagaaggtgaactagtagctcgtcttatttttagataaaagtaggtaaaaaaa
 ccaatcaagcgtttgtggatttagttcggttacaaaaaaagccatacaagaaggatttaaa
 40 cagttaaaaaatggaaacgcgtatcaaaatttatatgaagcagcacttgacgaagtga
 gcagattgtagtaggattaatgcaacacgtgcactaacgacaaaatatgatgcacaa
 ttatcattaggctcgtgtacaaaactccaacaatacaaatagttaaatcaagacaagatgag
 attaactatttttaaccagaaaaatattacacggttatccattaatgttgatggttacgat
 ttaaacccttaagcaacaaaagcgatataaagataaaaaagaattagaattgattgaacat
 aaaattaaacatcaagaaggaaagatattagaagttaaaggaaaaaataagaaatcttac
 45 gcgcaacctttattttaatttaacagatttacaacaagaggcatataaacattacaagatg
 gggcaaaggagacactaaatacattacaacatttatatgagagacataagttagtaacc
 tatccccgtacagattctaattattttaacagatgatatggtcgatacaattcaagaacg
 ttaagagcaatttttagctacagattataaatctcatgttcgagatttaatttctgagtc
 ttttcttctaaaatgcatatttttaataatcaaaaagtttcagatcatcatgcgattatt
 50 cccacagagggttagaccatctattgaacaattgagtcacgagagttlaaaatttatatg
 ctaactcgagaaagatttttagaaaatttaataatccttatttatatgaagttttaaca
 atccatgcacaactgaaagattacaattttgttttaaaagagataatacctaaacaatta
 ggatataaagctttaaaagatcaaacctcttcgcatactttaacgcattcttttaagaa
 ggtcagttattttaagttacatcgatttagattcatgaacatgaaacaaaggcaccggaa
 55 tatttttaacgaaggttcattacttaaaagccatggagaatccacaaaatcatattgatttg
 aatgataaaaagttatgcaaaaacactcaaacattcgggggggattggaactgtagcaact
 agggtgatatttatagaaaagttatttaacatgaatgcttttagagtcgcgagatggcaaa
 attaaagttacatcaaaaaggaaaacaaattttagaattgtctccaagtgaattaacctca
 cctataactaacagcccaatgggaagaaaaattaatgcttatcgaaaaggggaaatataat

tctcagaaattcatcacaggaaatgaaaaactttacatttaaagtagtaataaaattaaa
agcagtgagcaaaaatataaacatgataatttaacaacaaccgagtgccaacatgtggt
aagtttatgataaaagtcaaaactaaaaatggacagatgcttgtatgtcaagatcccaaa
tgtaaaactaagaaaaatattcaacgcaagactaatgcacgttgcccttattgtaagaaa
5 aaaaatgacttttattcggtaaagggaagaagctgtttatagatgtgtatgtggccacaca
gaaactcaatcacaaatggacaaaagaatgagagataaaaacgaatggtaaagtttcacgt
aaagaaatgaaaaatatataaataaaaaagaagaatcgacaataatccattcaaagat
gctctgaaaaatctcaaattgtag

10 Sequence 198
VTNATPEQYNPSYKEWNLEDLPPIPKMKMTVVISKTNRQFKIVKSLILDKNVKEIIATD
AGREGELVARLILDKVGNKKPIKRLWISSVTKKAIQEGFKQLKNGNAYONLYEAAALARSE
ADWIVGINATRALTTKYDAQLSLGRVQPTPTIQIVKSRQDEINYFKPEKYTLSINVDGYD
LNLKQQKRYKDKKELELIEHKIKHQEGKILEVKGKNKKSQAQPLFNLTDLQQEAYKHYKM
15 GPKETLNTLQHLIERHKLVTYPRDSNYLTDDMVDTIQERLRAILATDYKSHVRDLISES
FSSKMHIFNNQKVDHHAIIPTVRSIEQLSQREFKIYMLIAERFLENLMNPPLYEVLT
IHAQLKDYNFVLKEIIPKQLGYKALKDQTSSHTLTHSFKEGQLFKVHRIEIEHEHETKAPE
YFNEGSLLKAMENPNHIDLNDKKYAKTLKHSGGIGTVATRADIIEKLFNMNALESRDGK
IKVTSKGGKILELSPSELTSPILTAQWEEKMLIEKGKYNQKFIQEMKNFTFKVVKIK
20 SSEQKYKHDNLTTECPTCGKFMIVKTKNGQMLVCQDPKCKTKKNIQRKTNARCPYCKK
KMTLFGKGKEAVYRCVCGHTTETQSQMDKMRDKTNGKVSREMKKYINKKEIDNNPFKD
ALKNLKL*

Sequence 199
25 Contig_0456_pos_6881_6549,
is similar to (with p-value 1.0e-19)
>gp:gp|AE001272|AE001272_20 Lactococcus lactis DPC3147 plasm
id pMRC01, complete plasmid sequence. NID: g3582195.
atggctttttttgtaccgaactaatccacaaacgcttgattggttttttattacctact
30 ttatctaaaataagacgagctactagttcaccttctcgtccagcatctgttgcataata
atttctttaacatttttattctaaaattaaagattttacaattttaaattgtctatttgtt
ttactaatcactactgttttcattttcttaggaatgataggtaagtcttctaaattccat
tctttatatgaagggttatattgttcaggcgttgcatttgtcacaagatgccccaatgcc
caagttactatatactgtttcccttctatataa
35

Sequence 200
MAFFVTELIHKRLIGFLLPTLSKIRRATSSPSRPASVAIIISLTFLSKIKDFTILNCLFV
40 LLITTVFIFLGMIGKSSKFHSLYEGLYCSGVAFVTRCPNAQVTIYCFPSI*

Sequence 201
Contig_0456_pos_5929_5312,
is similar to (with p-value 4.0e-77)
45 >gp:gp|Y14043|SXY14043_1 Staphylococcus xylosus gltA, gdh ge
nes. NID: g2226000.
gtgcatttaaatagggtgtatctaaaaacgatgcctattttcaacgggtatgcaacttgctcgtt
accactctatttagcgctattttcttaggtgaatggagcacgattgttcaagtagtgatg
ggacttatagcaatgatcttattggttgtaggtatttctttaacatcacttaaagccaaa
50 agcgaaggcaaatccgataaaccagaatttaaaaaagcaatgggaatatcttctatca
acaatcggttacgtaggttatgtcgttcttgagatattttgtagtaagtggtagacagat
gctctcttctccaatcaattggtatggcaattggaggattaatcctttcaatgaatcat
aatacttcaattaaatcactgctctaaatcttataaccagggttatcttggggtatcgtt
aacttatttatgttctattcacaacctaaagttggtgtagcaactagtttctcattatca
55 caactgcttggtattgtttcaactttagggggtatctttattctaggggagaaaaagat
cgctcgccaaatgattggtatttggtcagggtattatcggttatagttatagcttcaatcatt
ttaggcaacttaaaatag

Sequence 202

VHLIGVSKTMPISTGMQLVGTTLFSAIFLGEWSTIVQVVMGLIAMILLVVGISLTSLKAK
 SEGKSDNPEFKKAMGILLSTIGYVGYVVLGDIFGVSGTDALFFQSIGMAIGGLILSMNH
 NTSIKSTALNLIPIGVIGIGNLFMFYSQPKVGVATSFSLSQLLVIVSTLGGIFILGEKKD
 RRQMIGIWSGIIVIVIASIILGNLK*

5

Sequence 203

Contig_0456_pos_5284_4493,

is similar to (with p-value 0.0e+00)

>gp:gp|Y14043|SXY14043_2 Staphylococcus xylosus gltA, gdh ge
 nes. NID: g2226000.

10

gtgtttgaagaattagaaaataaagtgggtcttattactggagctgccactggaattggc
 aaatctattgcggaaaatttttggtaaagctaaggccaaggtgtgtataaattaccgttct
 gatcgacatcatgatgaaattgaggaaattaaacaaactgttgctaaatttggtggtcaa
 acattggtggttcaaggtgatgtttcaattgaagaagatattaaacgaatgattgaaaca
 acaattaatcactttggaacttttagacattataattaataatgctggattcgaaaattca
 atcccaactcatgaaatgtcgattgacgactggcaaaaagttattgacataaacttaact
 ggcgcctttgtgggttcaagagaagccatcaatcaatttttaaaggaaaaacaagaaaggt
 actattattaacatttcgagtggttcattgacactattccatggcctaattatgtacactat
 gccgaagtaaggtggcttaaaattaatgatggaacaatgtcaatggaatatgcccaa
 tacggtattcgtattataatatatctcctggggcaattgttactgaacacactgaagaa
 aaattttctgacccaacgacgcgtgaagaaacaataaaaaatgatacctgcacgtgaaatt
 ggaaatgctcaagatgtagctaattgcagtactattcctatcttcagatcttgcaagttat
 atacacggtacaacattgtacggtgatgggtggcatgatgaactatccagcatttatgggt
 ggtaaaggttaa

25

Sequence 204

VFEELNKVVLITGAATGIGKSIAENFGKAKAKVVINYRSDRHHDEIEEIKQTVAKFGGQ
 TLVVQGDVSIIEEDIKRMIEETTINHFGTLDIIINNAGFENSIPTHEMSIDWQKVIDINLT
 GAFVGSREAINQFLKENKKGTTIINISSVHDTIPWPNYVHYAASKGGLKLMETMSMEYAO
 YGIRINNISPGAIVTEHTEEFSDPTTREETIKMIPAREIGNAQDVANAVLFLSSDLASY
 IHGTTLYVDGGMMNYPAFMGGKG*

30

35 Sequence 205

Contig_0456_pos_2714_1809,

putative peptide of unknown function

atgagtagtactcgtaaacaaaattagattatgaggaacaaattaaaaagttgaaatca
 ttaggaattctattcaatgaaataacagaagaagactaaagaaatattaaaaataac
 acttatttttttaattgatattcttttcgtaaaaataaaaaaggatagtagtggaat
 tataatttttgagttttctgcactttctgattttgctacttttagatatgcgattaagatat
 actttattacctatgtgtttggatatagaacattcactaaaaacagatatcttataaaag
 attactgatgatgtaaacgaagacggatatacaattgttcaagattttataaacaatcat
 aatggagatttagaaaaaatcttttctagcgtgattaaaaagagatggtacagttataccg
 agttttcaaaaaatattatgatgatcctccaatatgggtatgcttagaattaatgactttt
 ggccaattttcagcattttgtagaattttattctgaaagaacaaatgactctgagttacgt
 aaggctggtaaaatttattaaatttgctaaaaacattaggaataaatgtgctcatagccaa
 ccaatttttattaaatttaaattccacgcaaaaactttaccggtgaaagagaattaaaaaag
 ataggtagaaaaacaaagactgtctgataaaaaaccttaaagatttagcaataattgatatt
 ctgcatattattagttttacattctaaatattgtagtaaaaggtataaaagataatcgaaaa
 aatgattttattaacttttaaacacgtaaaaaatagatattttcatcattatcgaaatgtt
 ccttctctttcatttttctttctatcacttaacaaaatgattgactattatgttcaaaac
 aattaa

55

Sequence 206

MSSTRKPKLDYEEQIKKLKSLGILFNEITEEEAKEILKNNTYFFKLISFRKNIKKDSSGN
 YNFEFSALSDFATLDMRLRYTLPMCLDIEHSLKTDILKKITDDVNEDGYTIVQDFINNH
 NGDLEKIFSSVIKRDGTVIPSFQKYDDPPIWVCLELMTFGQFSAFVEFYERTNDSELR
 KAGKFIKFAKNIRNKCAHSQPILLNLNPRKNFTVERELKKIGRKQRLSDKNLKVLAII DI

LALLVLHISKYCSKGIKONRKNDDLTFKQRKNRYFHRYRNVPSLSFFFLSLNKMIDYYVQN
N*

Sequence 207

5 Contig_0456_pos_0_539,
is similar to (with p-value 3.0e-49)
>sp:sp|P39755|NDHF_BACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.
6.5.3) (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5). >gp:gp|U283
23|BSU28323_1 Bacillus subtilis NADH dehydrogenase subunit 5
10 (ndhF) gene, complete cds. NID: g903586. >gp:gp|Z99104|BSUB
0001_183 Bacillus subtilis complete genome (section 1 of 21)
: from 1 to 213080. NID: g2632267. >gp:gp|Z99105|BSUB0002_11
Bacillus subtilis complete genome (section 2 of 21): from 1
94651 to 415810. NID: g2632457. >gp:gp|AB006424|AB006424_9 B
15 acillus subtilis genomic DNA, 70 kb region between 17 and 23
degree. NID: g3599592.
atgcattaccgtaaataattttccggtttttacattaattactgcatttgcttcaattggca
tggttaagtggagacttaagggttaatgacatggtttgggtgcaacattattgtgtta
acacggctcattaaagttaacaaattatggaaggtgcctaggaagcagcaagaatttca
20 gcttggtcatttatattggcatggttgcgttattgattgctgtcattttattgtatatac
gctacaggagattggtatattttattcgaatatgtcagatgataatgcaatcaattatgga
atgcgtctctgtatcaatttacttattgttttagctgtgattattccggcggcacaattt
ccatttcaaggctggcttattgaatctgtagctgcgcctacgccagtttcagctattatg
cacgctggtattgttaatgctggtggcgttattcttacacgcttttctccggtattta
25 gacgaaatagccatttctactgtttattaattattgcaagtatcttcagattgttgggttc

Sequence 208

MHYRKYFPFFFTLITAFASLAWLSGDLRLMTMFWGATLFVLTRLIKVNKLWKVPREARIS
AWSFILAWLSLLIIVILLIYIATGDWYIYNMSDDNAINYGMRLCINLLIVLAVIIPAAQF
30 PFQGWLIESVAAPTTPVSAIMHAGIVNAGGVILTRFSPVFNDEIAISLLLIASISVLLGS

Sequence 209

35 Contig_0457_pos_1064_2419,
is similar to (with p-value 2.0e-79)
>sp:sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENS
OR PROTEIN PHOR (EC 2.7.3.-). >pir:pir|A27650|A27650 regulat
ory protein phoR - Bacillus subtilis >gp:gp|AF008220|AF00822
0_180 Bacillus subtilis rrnB-dnaB genomic region. NID: g2293
40 135. >gp:gp|M23549|BACPHORP_2 Bacillus subtilis alkaline pho
sphatase regulatory protein (phoP gene, 3' end and phoR gene
, complete cds). NID: g143329. >gp:gp|Z99118|BSUB0015_175 Ba
cillus subtilis complete genome (section 15 of 21): from 279
5131 to 3013540. NID: g2635200.
45 atgcgttacacctataagaatacaatagatgataaaacaatatacataagtggaaattaat
aatgaaattattgattttacaaaaagatttatggaaatacttgtctattgttgagtcatt
gtattattttacggtttatttagcaagtagaagtagatcaatcgaacatatattagacctatc
aatgaagtaacttatgctacatcacttctagcagatggatattaccatgttcgtgttcca
gaaagtaattgtgaaggaaactagggcattatttgtgactacaaatgacttagcacgacga
50 ttgcaaaaattaaacaatagtcaaaaaattcaatccaatagattaaaaaactaccttagaa
aatataccgagttcagtagtattgataaacatggagaaattgtagttgctaatacat
gcttattatcaggtgtttaaccctgatcaaatggtagaaaaataaaagttacattgggttc
atagatgattgattgaaaaatttaattattgaaagtttagaactgaaaaagttatctat
gaacaattagaagttgctattaataacgtacataactaaatatttcgatgtatcttgcatac
55 cccatttttaactaaatctaaaaaaatttacaaggatggtggttgcattcatgacatt
actaatttgcagaaattagaaaaccttagaagggaatttgttgcaaatgtgtcacatgaa
ctaaaaacaccgattacttcaatcaaagggttttgcaaaactctgattgaagggtgctaaa
aatgatgaacaatcgcttgatattgttttaaatatttttaaaagaatctaatagaata
gagtcattggttacagacttatttagatttatcacatatagaacagcaaaaagaacttgaa

ataaattacatgaatttatctgaattagctattaatataatagataatttgcaaacacaa
gcatacaataagagaatcaaaatacaatctgaaattgaaaaagatgtcatcattgaggca
catgaaaataaaatagcgcaagttattactaatttgctatcaaattgctataaattattct
tcagaagataataaggtaatagtaagagtatatagaaatgacaataaagtttatttagag
5 attcaagattatggtattggtataaagtgaacagatcaaaagcgtatatttgaacgtttc
tatcgtgtagataaaagcgagaagtagagattcaggtggtacaggacttggctctgtctata
acaaaacatatgttgaagcacataatggtagaatagacgtgaaaagtgcacctggcaaa
ggttcgatattcaaagttctatttaataatgataattaa

10 Sequence 210
MRYTYKNTIDDKTIYISGINNEIIDLQKDLWKYLSIVGVIVLFTVYLASRSINRTYIRPI
NEVYATSLLDGYHVRVPESNVKETRALFVTTNDLARRLQKLNNSQKIQSNRLKTTLE
NIPSSVLMIDKHGEIVVANHAYYQVFNPDQMVENKSYIGFIDDSIEKLIIESFRTEKVIY
EQLEVAINNVHTKYFDVSCIPILTKSKKNLQGMVVVLHDITNLQKLENLRREFVANVSHE
15 LKTPITSIKGFAETLIEGAKNDEQSLDMFLNIILKESNRIESLVTDLDSLHIEQQKELE
INYMNLSELAINIIDNLQTQAYNKRIKIQSEIEKDVIIEAHENKIAQVITNLLSNAINYS
SEDNKVIVRVYRNDNKVYLEIQDYGIGISETDQKRIFERFYRVDKARSRDSSGGTGLGLSI
TKHIVEAHNGRIDVKSAPGKGSIFKVLFDND*

20 Sequence 211
Contig_0457_pos_3248_4048,
is similar to (with p-value 2.0e-37)
>sp:sp|P13252|DPO1_STRPN DNA POLYMERASE I (EC 2.7.7.7) (POL
I). >pir:pir|A32949|A32949 DNA-directed DNA polymerase (EC 2
25 .7.7.7) - Streptococcus pneumoniae >gp:gp|J04479|STRPOLA_1 S
.pneumoniae DNA polymerase I (polA) gene, complete cds. NID:
g153764.
atgaaaggtctaattgggggatacctctgacaatattcctggcggttgcgtggtgcggcgaa
aagacggctattaaattacttaataatcaatttgagtcagtagaaggggtctatgaacatatt
30 gaggaggtcactgcaaaaaaattaaaagaaaaactcatcaatagtaaagatgatgcctta
atgagtaaagatttagcaacaatcaatgttcacagtcaggattgaagtatcattagaagat
acaaaattaactctacaagacgacactacagaaaaaattgaactatttaaaaagctagaa
tttaaaccaactattagcagatatagacacatcctctacgaatgaagaagtcataagataaa
acttttgaaattgagcaagactttcaaaaatgtagatttgaatgatttaaacgaagcggta
35 atacattttgaactcgaaggcactaattatcttaaaagacactattctcaagtttggtttt
tatacaaatcatcaacatgtagtgataaatgctgaggatgtaaaggattataaacattta
gttcaatggccttgaagataaaaaatacaactaaaattgtctatgatgcaaaaaaaacttat
gtatctgctcatcgattagggattaatatagaaaatattgaatttgatggtatgtagca
agctatattattgacccatcacgttctattgatgacgttaaactctgtggttaagtttatat
40 ggacaaaattatgtaaaagataatattacaatatttgggaaaggttaagaaacatcatata
cctgaatatccctcattttaa

Sequence 212
MKGLMGDTSDNIPGVAGVGEKTAIKLLNQFESVEGVYEHIEEVTAKKLKEKLINSKDDAL
45 MSKDLATINVHSPIEVSLEDTKLTLODDTTEKIELFKKLEFKQLLADIDTSSTNEEVIDK
TFEIEQDFQNVLDLNEAVIHFELEGNTYLNKDTILKFGFYTNHQHVINAEDVKDYKHL
VQWLEDKNTTKIVYDAKKTYVSAHRLGINIENIEFDVMLASYIIOPSRSIDDVKS SVSLY
GQNYVKDNITIFGKGKKHHIPEYPSF*

50 Sequence 213
Contig_0457_pos_4381_5253,
is similar to (with p-value 0.0e+00)
>gp:gp|U02682|HIU02682_1 Haemophilus influenzae KW20 catalas
e (hktE) gene, complete cds. NID: g409459.
55 gtgattcctgaaagcgtcgatgcagcgaaaggttcaggtgcatttggtacgttcacagtt
acaaatgacatcacacaatatatacaaatgcgaaaatattctcagaagtcggaaaacaaaca
gagatgtttgacaggttttttactgtttcaggagaacgtggagcagcagatttagaacgt
gatatacgtgggtttgccttgaaattctacactgaagatggaaactgggatttagtaggt
aacaatacgcaggttttcttcttttagagatcctaaactatttattagtttgaatcgtgct

gtaaaacgagatccacgtacaaatatgagaagtgacaaaaataactgggacttttgaca
 ggtctaccggaagcattgcatcaagtgacaatattaatgtcagatagaggtagtgcacaaa
 ggattccgaaatatgcatggattcgggttctcatacgtattctatgtataatgataaagg
 5 gaacgtgtatgggtaaaaatcatttccgtacacacaagaattgaaaactatactgac
 gaggaagcagctaaaaattgtaggtatggatagagattcttcacagagggatttatataat
 gctatcgaaaatggagattatccaaaatggaaaatgtacattcaagttatgacagaggaa
 caagctaaaaatcatccagacaatccttttgatttaacaaaggtaggtataaaaaagac
 tatccactgattgaagtgggagaatttgaattgaatcgtaatcctgagaattattttctt
 gatgtagagcaggcagcggtttacgcctacaaatatgttctctgggttagattattcacca
 10 gataaaatgctacaaggacgtttattctcataa

Sequence 214

VIPERRMHAKGSGAFGTFVTNDITQYTNAKIFSEVGKQTEMFARFSTVSGERGAADLER
 DIRGFALKFYTEDGNWDLVGNNTPVFFFRDPKLFISLNRAVKRDPRTNMRSAQNNWDFWT
 15 GLPEALHQVTILMSDRGMPKGFERNMHGFGSHTYSMYNDKGERVWVKYHFTQQGIENYTD
 EAAKIVGMDRDSQRDLNNAIENGDPKWKMYIQVMTEEQAKNHPDNPFDLTKVWYKKD
 YPLIEVGFEFLNRNPENYFLDVEQAAFTPTNIVPGLDYSPDKMLQGRIFS*

Sequence 215

Contig_0457_pos_6680_5622,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF090142|AF090142_1 Staphylococcus epidermidis lipase
 precursor (gehD) gene, complete cds. NID: g3789931.
 gtgatttttttgaaaaataataatgaacaagaagatttagcattaggaagtacacggtg
 25 ggagtcgtgtcaatcattactgggattacaatatttgcagtggtcagcatgctcaagct
 gctgaaatgacacaatcatcatcagattttaacgaacagtcacaacaacagaacaagtt
 gaacacaaagaagatacaactcatttatcatcacgaattgaatcaagagggtgacacagct
 agccaatcaaagactaatcaagagaaccaatctgatgaaaatgtacaaaaaagaataat
 caaactcaacaagattcaacacaaacgtcaccattaaatgaccaagaacaaactttaag
 30 gggcaacaatcaaaagacaatcatgttaccccaaattcacgtcaggatacatatccaaaa
 ggccaaaatcaagatgataaaggcaacaacagtttaagataatcaacactcacaaaca
 gaacatcaacctaataactcaaaacaaaataatgatcaagattcatcagataaaaagcaa
 caccatctgatcaaaactcaagcccatcttcaaaaggacacaacctaaacaatcacag
 tctataggagatagataaaaacagtaaaaacaacctcttctaaagtacacaaaataggt
 35 aatacaaaaaactgataaaaacagttaaaacaaatcaaaaaagcaaacatcattaacttca
 ccacgcgttgtgaaatcaaaacaaactaaacatatcaatcaacttactgcgcaagctcaa
 tataaaaatcaatatccagtcgtgtttgtacatggattttaggttagtcggtgaagat
 tcatcagcatgtacccaattattgggtggtactaaatataacgtgaacaagaactt
 acaaaattaggttaccgagttcacgaagccaatgtaggagcatttagcagcgggtgaagtt
 40 aatttacgattggaccgcgtgtaccatttcaattcgtcttggtcaaaaattcatcgtcg
 tacttagtattatcactgatatttctcatgtttatcataa

Sequence 216

VIFLKNNNETRRFSIRKYTVGVVSIITGITIFVSGQHAQAEMTQSSSDFNEQSQQTEQV
 45 EHKEDTTHLSYELNQEEDTASQSKTNQENQSDENVQKNNQTQQDSTQTSPLNDQEQLK
 GQSKDNHVTNPSRQDTYPKGQNDKKGKQFQKDNQHSQTEHQPNQNTQNNNDQSSDKKQ
 HPSDQQTAPSSKGTQPKQSISIGDRDKTVKQPSSKVKHIGNTKTDKTVKTNQKKQTSLS
 PRVVKSKQTKHINQLTAQAQYKNQYPVVFVHGFVGLVGEDSFSMPYNYWGGTKYNVKQEL
 TKLGYRVHEANVGAFSSGEVNLRLDRVYHFNSSWLKNSSSYLVLSLIFSCLS*

Sequence 217

Contig_0458_pos_6103_5078,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P49787|ACCC_BACSU BIOTIN CARBOXYLASE (EC 6.3.4.14) (A
 55 SUBUNIT OF ACETYL-COA CARBOXYLASE (EC 6.4.1.2)) (ACC). >gp:
 gp|U36245|BSU36245_2 Bacillus subtilis biotin carboxyl carri
 er protein (accB) and biotin carboxylase (accC) genes, compl
 ete cds. NID: g1055244.
 atgggaataaaagatattgctaaagctgaaatgattaaagccaatgtacctgtagtacca

ggaagtgaaggacttattcaaagtatagatgacgctaaaaaatagctaaaaaatcggc
 tatccagttatcatcaaagccacagcaggtggtggtggaaggtattcgggttgctcgt
 gatgagaaagaacttgaaactggttaccgtatgacacaacaagaagctgaaaccgcgttc
 5 ggaatggtggtttatacttagaaaaatttatagaaaactttagacatatagagattcaa
 attattggcgatacttatgaaacggttatacatttaggtgaacgtgattgtacaattcaa
 agaagaatgcaaaagctcggtgaagaagcaccctcaccagttttaagtgaagataaacgc
 caagaaatgggtaatgctgcaattagagccgcaaaagctgtaaattatgaaaacgcaggt
 acaattgaatttatatatgatttagatgataaccaattttatttcatggaaatgaataca
 10 cgtattcaagttgaacacccagtaactgaaatggtaacaggagtagatttagtaaaatta
 caactcaaagttgctatgggtgaggcgttaccttttaacaagaagatatttccattaac
 ggtcacgctattgaatttcgaatcaatgctgaaaatccttacaaaaactttatgccatca
 ccaggcaagattacccaatatcttgctccaggcgggttttgagtgagaattgaatcagca
 tgttatactaatttatacgataaccaccttactatgactccatggtggcaaaacttatagtt
 cacgaacctacacgtgaagaatcaattatgacaggcattcgtgctttaagtgaatatctt
 15 gttttaggtatcgacactacgattccattccacttaagacttctaataatcatatTTTT
 agaagtggggaatttaatacaaaattcctagaaaagtataatattatggacgataataac
 caatag

Sequence 218
 20 MGIKDIAKAEMIKANVPVPGSEGLIQSIDDAKKIAKKIGYPVVIKATAGGGGKGIRVAR
 DEKELETGYRMTQQEAETAFNGGLYLEKFENFRHIEIQIIGDTYGNVIHLGERDCTIQ
 RRMQKLVEEAPSPVLSEDKRQEMGNAAIRAAKAVNYENAGTIEFIYDLDDNQFYFMEFNT
 RIQVEHPVTEMTGVLDLVLQKLVAMGEALPFQEDISINGHAIEFRINAENPYKNFMP
 PGKITQYLAPGGFGVRIESACYTNYTIPPYYDSMVAKLIVHEPTREESIMTGIRALSEYL
 25 VLGIDTTIPFHLRLNNHIFRSGEFNTKFLEKYNIMDDNNQ*

Sequence 219
 Contig_0458_pos_5066_4704,
 is similar to (with p-value 3.0e-18)
 30 >sp:sp|P54519|YQHY_BACSU HYPOTHETICAL 14.7 KD PROTEIN IN ACC
 C-FOLD INTERGENIC REGION. >gp:gp|D84432|BACJH642_218 Bacillu
 s subtilis DNA, 283 Kb region containing skin element. NID:
 g2627063. >gp:gp|299116|BSUB0013_144 Bacillus subtilis compl
 ete genome (section 13 of 21): from 2395261 to 2613730. NID:
 35 g2634723.
 atggtcaatgtagcagattattctcaatctaatttaggaaaaattgaaatagcaccagaa
 gtattatctgttatcgcatccattgcgacatcagaagtagaaggtattacaggccatttt
 gctgaactaaaaaaaaacaatctagagaagattagtcgaaaaatttaaacagagattta
 aaaatcgaagctaaagaagacggaatatacattgatgtattttgttctttaaacatggc
 40 gtaaatattttctaaaactgcaaatcaaatcgaagcaattttcaattcaattacgaca
 atgacagctattgaaccacagcaaatatattcacatcagaagtatcgtcgcagaaaaa
 taa

Sequence 220
 45 MVNVADYSQSNLGKIEIAPEVLSVIASITSEVEGITGHFAELKKTNLEKISRKNLNRDL
 KIEAKEDGIYIDVFCSLKHGVNISKNTANQIQEAFNSITMTAIEPQQINIHRSIVA
 EK*

Sequence 221
 50 Contig_0458_pos_4628_4239,
 is similar to (with p-value 2.0e-23)
 >sp:sp|P54520|NUSB_BACSU N UTILIZATION SUBSTANCE PROTEIN B H
 OMOLOG (NUSB PROTEIN). >gp:gp|D84432|BACJH642_219 Bacillus s
 ubtilis DNA, 283 Kb region containing skin element. NID: g26
 55 27063. >gp:gp|299116|BSUB0013_143 Bacillus subtilis complete
 genome (section 13 of 21): from 2395261 to 2613730. NID: g2
 634723.
 atgagtcgtaaagatgcaagagtacaagcttttcaaactttatttcaacttgaaataaaa
 gagacagatttaacaattcaagaagcaattgaatttattaaagatgatcattctgattta

gactttgattttatatactggtagttactggagtc aaagatcatcaa atcgtttagac
 gaaacaattaaccccat ttaaaagactgggtctatcgatcg tttactgaaatcagatcg t
 attattttaagaatggcaacttttgaaatattgcacagcgacacaccta aaaaagtagtt
 gttaatgaagctgtagaactc aaaaacagtttagtgatgatgatcattataaatttgtt
 5 aatgggtgttttaagtaataataatgattaa

Sequence 222
 MSRKDARVQAFQTLFQLEIKETDLTIQEAEFIKDDHSDLD FDFIYWLVTGVKDHQIVLD
 ETIKPHLKDWSIDRLKSDRIILRMATFEILHSDTPKKVVVNEAVELTKQFSDDDH YKFV
 10 NGVLSNIND*

Sequence 223
 Contig_0458_pos_2587_1805,
 is similar to (with p-value 3.0e-69)
 15 >sp:sp|Q08291|ISPA_BACST GERANYLTRANSTRANSFERASE (EC 2.5.1.1
 0) (FARNESYL-DIPHOSPHATE SYNTHASE) (FPP SYNTHASE). >pir:pir|
 JX0257|JX0257 geranyltranstransferase (EC 2.5.1.10) - Bacill
 us stearothermophilus >gp:gp|D13293|BACFDPS_1 B. stearotherm
 ophilus DNA for farnesyl diphosphate synthase, complete cds.
 20 NID: g391609.
 atgaaatatttcattaaatgctgggtggttaaagaatcagaccagtcattattattattaaca
 ctaaaaatgcttaacaaagattatcaacaaggactaaatagtgcttttagcattggaaatg
 attcatacttattctttaattcatgatgatttaccagcaatggataatgacgattaccgt
 agaggaaaattaacaaatcataaagtttatggatgaatggaaagccattcttgctgggtgat
 25 gcattattaacaaaagcttttgaattagtttctaataatgatactaccattgaagatagtggtg
 aaagtaagtattataaaaagactttcaaaaagcaagtggacatttgggaatgggtgggtggc
 caagcgcttgatatggaaagtgaaggggaagtcaattcgtttagaaactttagaatcaatt
 catgaaactaagacagcgctttactaaatttttcagttatggctgcggtagacattgct
 caagtagaacaataattgctaagaatttagatgaatttagtcatcatttaggaatgatg
 30 tttcaaatataagatgatttactggatgtgtatgggtgatgaatcaaaacttggcaaaaaa
 gtaggcagtgatataagtaaatcataaaagtacttatgtttctttacttggaaaagaagga
 gcagaagaaaagttaaacaatcatcaatatcttgctatgaactgcttaaatcaaatttct
 gatcaatatgatacttctgaattaaagtgatattgtagatttattctataacagagaccat
 taa

Sequence 224
 MKYSLNAGGKRIRPVILLTLKMLNKDYQQLNSALALEMIHTYSLIHDDL PAMDNDYR
 RGKLTNHKVYGEWKAILAGDALLTKAFELVSNDDTTIEDSVKVSIIKRLSKASGHLGMVGG
 QALDMESEGKSIRLETLESIHETKTGALLNFSVMAAVDIAQVEQNI AKNLDEFSSHLLGMM
 40 FQIKDDLLDVYGDSESKLGKKGVSIDVNHKSTYVSLGKEGAEEKLNHQLAMNCLNQIS
 DQYDTSELSDIVDLFYNRDH*

Sequence 225
 Contig_0458_pos_0_1022,
 is similar to (with p-value 1.0e-74)
 45 >sp:sp|P17894|REC_N_BACSU DNA REPAIR PROTEIN REC_N (RECOMBINAT
 ION PROTEIN N). >pir:pir|B35128|B35128 recN homolog - Bacill
 us subtilis >gp:gp|D84432|BACJH642_227 Bacillus subtilis DNA
 , 283 Kb region containing skin element. NID: g2627063. >gp:
 50 gp|M30297|BACREC_N_2 B. subtilis recombination and sporulation
 protein (recN, spoIVB) genes , complete cds, arginine hydro
 ximate resistance (ahrC) gene, 3' end. NID: g143400. >gp:gp|
 Z99116|BSUB0013_135 Bacillus subtilis complete genome (secti
 on 13 of 21) from 2395261 to 2613730. NID: g2634723.
 55 atgttacaaaccttatcaataaaacaatttggcattattgacgaacttgatataaaacttt
 tctgacggtctaacagttatgagtggtgaaactggctcaggaaaatctatcattattgat
 gccattggacagttaatcggtatgagagcttcttctgattacgtcagacatgggtgaaaag
 aaagcaattatcgaaaggtatctttgatatagacgagagtaaagacgcaattaatatacta
 gaatcatttagctatagatgttgatgaagattttttattagttaaaagagaaattttcagt

tctggtaagagtatttgcgtattaataaccaaactgtcactctacaggacttaagaaaa
 gtgatgcaagaactgcttgatattcatgggtcaacatgaaacgcaatctttacttaagcaa
 aaatatcatcttcaactattagatgattatgcagacaatcagatttcagatttacttaat
 caatatcaactttcttataaccaatataaaaaataaacgtaaagaattagaggaattagaa
 5 tccgcggaccaggctttattacaacgatttagacttaatgaaatttcaattagaggaacta
 accgaagcttcactgaaagaaggcggaagtggaccaacttgaatccgatattaaaagaatt
 caaaactccgaaaaattaaatctagctttaaacaatgcacatcaagttctaactgatgaa
 agtgcaatacccgataggttgtaacgaattaagcaactacttgcaaacgattaatgatatc
 gttccagaaaaattcgttaagattaaaagaggacattgatcaattttactatatgctagaa
 10 gatgcaaaagcatgaaatttacgacgaaatggctaacactgaattcgatgagcaagtttta
 aatgagtatgaatccagaatgaatttacttaataatttaaacgtaaatatggtaaggat
 attactgaacttattgcttatcagagtaaaccttgcaaatgaaattgataaaatagTGGAA
 TT

15 Sequence 226

MLQTLISIKQFAIIDELDINFSDGLTVMSETGSGKSIIDAIQQLIGMRASSDYVRHGEK
 KAIIEGIFDIDESKDAINILESIAIDVDEDLLVKREIFSSGKSICRINNQTVTQLQDLRK
 VMQELLDIHQGHETQSLLKQKYLQLLDDYADNQYSDLLNQYLSYNQYKNKRKELEELE
 SADQALLQRLDLMKFQLEELTEASLKEGEVDQLES DIKRIQNSEKLNALNNAHQVLTDE
 20 SAI PDRLYELSNYLQTIINDIVPEKFVRLKEDIDQFYMLEDAKHEIYDEMANTEFDEQVL
 NEYESRMNLLNNLKRKYGKDITELIAYQSKLANEIDKIVEX

Sequence 227

Contig_0459_pos_802_1155,
 25 putative peptide of unknown function
 atgaaattcttaaatataaaattctctagctggtagctttttcttcaaactcctgctaaa
 aattcttcattatttgcacgctaatttcactctctctactgaaattcctttttcattaacg
 gtaataataccatatacgggtgttgaaccattgttcaaacctactgatccaggattaaaa
 tatactgttgatttatcatcaaacatatgcaacctatggttatgtccaaataaaattaaa
 30 tcggcttctttgtctttaataattcagaaatagcttggtcgtcatcttttgtaataggt
 gcaaaaggttggtcatcaataggagctgacattttatcattttcattttcataa

Sequence 228

35 MKFLNINSLAGTCFSSNPAKNSSLLYANFTLSTEIPFSLTVIIPYTVFEPLFKPTDPGLK
 YTVDLSSNICNLWLCPNKIKSASLSLNNSEIACSSSFVIGAKGCSSIGADILSFSIS*

Sequence 229

Contig_0459_pos_1809_2813,
 putative peptide of unknown function
 40 atggttagctcaactcggtagaatctactacaaaacctataatatctattatttttatttta
 ctcatcttagctttattattttttgttaggttaccattgataactggtacagtatat
 gcaattcaaaaagctattaataaagagaaagttcttttagtgatttatttttgctttt
 aaaaaaggcaaatatgctaaatcagtaatttttagctttaataacttttagttttattcatt
 gtaatcgtacttattctagtgtctattaataaaattatatagtttagctcttagcccaata
 45 ttaatcggttacaacaatcaataagcggctacgacaatccaatgggaattttaattaca
 atacaaattgtgtgttactcataacaggtttcatctcatcaattttctattggtttgta
 attatatctattattaattatactaccgcttatacagaagattcatctcgtaaagtaatg
 agtaatttaaaagaaggatttaaaaggtattaaaaacggtaaagaaaacttggtttaattt
 ttcattggcgtattacttattagtttacttgcaagtattattaacaaaccgctattatc
 50 ggtgtacaatacttaacaagcagtatgtctcaaacgggtggtcaaactattattataata
 gctagaatcgatatctatagattacgcctatgtctttattacattttgatttttggaatt
 attaatattttcggttagacgtggtgacaaaccagtcaaaaagcaaaagacgtcataaaaaat
 aaagatattaacaaaggtaatgtaaacgacaaagtagataactaaattaaatgcttccaac
 tccaaagatacagaagcagataaaatgaaagatcaacaaacacatatacaacaagacaaa
 55 actgatagtcagaaaaataacatatatgattccatcaaagaaaaagtaaatgaaaataaa
 gaaaatggttacagaacaatctaaaaatctatttgataagaaatag

Sequence 230

MLAQLGESTTKPIISIIFILLILALLFIFVGYPITGTVYAIQKAINKEKVLFSDLFFAF

KKGKYAKSVILALITLVLFIVIVLILVLLNKLYSLALSPILIGLQQSISGYDNPMGILIT
 IQIVLLLITGFISSIFYWVFIIFIINYTTAYTEDSSRKVMSNLKEGFKGIKNGKKTWFKF
 FIGVLLISLLASIINKPLLFGVQYLTSSMSQTVAQTIIIIARIVSIVLRLCLYYILIFGI
 INYFVRRGDKPVKSKRRHKNKDINKGNVNDKVDTKLNASNSKDEADKMKDQQTHIQQDK
 5 TDSQENNIYDSIKEKVNNENKENVTEQSKNLFDDK*

Sequence 231

Contig_0459_pos_5687_4470,

is similar to (with p-value 0.0e+00)

10 >sp:sp|P39754|GLMS_BACSU GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE A
 MINOTRANSFERASE (ISOMERIZING) (EC 2.6.1.16) (HEXOSEPHOSPHATE
 AMINOTRANSFERASE) (D-FRUCTOSE-6- PHOSPHATE AMIDOTRANSFERASE
) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERAS
 E) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
 15 atgttacaaactacaaaccaatacaaaagagatacatgacatgaaatagttattgttaag
 cgagacacagtagaaattaaagatcttgaggggcacattcaacaacgtgatacgtatacg
 gcagaaatagatgctgctgatgcagaaaaagcgatatgatcattacatgttaaaagaa
 attcatgaacagcctgcagtgatgcgtcgcatattcaagaatatcaagatgaaaaaggt
 aatttaaaaatcgattcagagattattaatgatgtagcagatgctgatcgtatttacatc
 20 gttgcagctggtagttatcatgctggattggttggttaaagaatttattgaaaaatgg
 gcaggtgtacactactgaggttcatgtagcttctgaatttgtatataatatgccacttctt
 tctgaaaaaccactattttatttatatttcacaatctggtgaaacagctgatagtcgtgct
 gtattagttgaaacaaataagtttaggtcacaatcattaacaattactaatgttgctggt
 tcaacattatcacgtgaagcggatcacattactttacatgctggacctgagattgca
 25 gtgcgcatctacaaaagcatatacagcgcaaatgctgttttatctatcttatctcaaatt
 gttgctaaaaatcatggtcgtgaaaccgatgttgatttattaagagaactagctaaggtt
 actacagctattgaaacaattgttgacgatgcacctaagatggagcaaattgcaacggat
 ttcttaaaaactactcgtaatgcattcttcattggacgaacaattgattataatgttagt
 ttagaaggtgcattaaaaattaaaagaatttcttatattcaagctgaaggatttgacggt
 30 ggggaattaaagcacggaacaatcgctttgattgaagatggcacacctgttataggttta
 gctacacaagaaaacgttaattctatcaattcgtggaaatatgaaagaagtagtagcacgt
 ggtgcatatccttgtagatttcaatggaaggttgaataaagaaggagacacatacgtg
 attccacaagtacatgaattattaactccttttagtatctgtagtgacaatgcaattaatc
 tcatattatgctgcgttacaacgagatttagatgttgacaaacctcgtaacttagccaaa
 35 tcgggttacagtagagtaa

Sequence 232

MLQTTNQYKEIHDEHIVIVKRDTVEIKDLEGHIQQRDTYTAEIDAADAEKGVYDHYMLKE
 IHEQPAVMRRIIQEYQDEKGNLKIDSEIINDVADADRIYIVAAGTSYHAGLVGKEFIEKW
 40 AGVPTEVHVASEFVYNMPLLSEKPLFIYISQGETADSRVLVETNKLGHKSLTITNVAG
 STLSREADHTLLHAGPEIAVASTKAYTAQIAVLISLSQIVAKNHGRETDVDLLRELAHV
 TTAIETIVDDAPKMEQIATDFLKTTRNAFFIGRTIDYNSLEGALKLKEISYIQAEFGAG
 GELKHGTIALIEDGTPVIGLATQENVNLSIRGNMKEVVARGAYPCMISMEGLNKEGDTYV
 IPQVHELLTPLVSVVTMQLISYYAALQRDLVDKPRNLAKSVTVE*
 45

Sequence 233

Contig_0459_pos_3987_3118,

is similar to (with p-value 3.0e-19)

50 >gp:gp|Z99122|BSUB0019_82 Bacillus subtilis complete genome
 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 >gp:gp|Z92954|BSZ92954_6 B.subtilis yws[A,B,C,D,E,F,G] and g
 erBC genes. NID: g1894764.
 atggaggataatatgaaaaatataaaagcaatttttttagatatggatggaacgatttta
 catgaaaaataataaagcctcggaatataactaaacaagtgattaatgaattgagagagcaa
 55 aattataaagtttctcttgctactggagatcttattcagaaatcagtcagcttgctcct
 gatggattcactgtagatggtattatcagttcgaatggaacttcagtgaaattcatgga
 gataatttgttttagacatagtttaacttttagaacgagtacagaaaattgtggaattggct
 aaaaaacaacatatttattatgaagtttttcttttgaaagtaatcgatatctcttaaa
 gaagatgaagattggatgaaagaatgatttccactatagagccacctgacgctgtaagt

caaagtgagtggtcatcggaagagaggcaattaaaggaaaaatagattggcgagataacc
 ttacctgatgcacacttttctaaaaatatatttatttagtcccaacttagataaaataact
 gattttcgcaaccagcttgggtgaaaaccaatcaaatttaggtattaccgtatctaattct
 tcgcggtataatgctgaaacgatgccatatcatcacagataagggtagaggtatcaaggaa
 5 atgattgatcactatggtattaagcaagaagaaactttagttattgggtgatagtataat
 gatagagctatgtttaattttggccatcacactgttgcaatgaaaaatgcaagacaagaa
 attaaaaatcttacagatgatattaccgaatacacgaacgaagaagatgggtgcagcacat
 tacttaaaaagtcatttattagataactag
 10 Sequence 234
 MEDNMKNIKAIFLDMDGTILHENNKASEYTKQVINELREQNYKVFLATGRSYSEISQLVP
 DGFTVDGISSNGTSGEIHGDNLFHSLTLERVQKIVELAKKQHIYYEVFPFESNRISLK
 EDEDWMKEMISTIEPPDAVSQSEWSSRREAIGKIDWRDTPDAHFSKIYLFSPNLDKIT
 DFRNQLVENQSNLGITVSNSSRYNAETMPYHTDKGTGIKEMIDHYGIKQEETLVIGSDSN
 15 DRAMFNFGHHTVAMKNARQEIKNLTDDITEYTNEEDGAAHYLKSHLLDN*

Sequence 235
 20 Contig_0459_pos_1574_798,
 putative peptide of unknown function
 atgttagaatatgctaataagataaacattgaaggagataaatatcatgaaatttgctggt
 atcactgatattcatggaaactttgatgcgcttcaaactgttttagatgatattgatagt
 agagatgatatcgaaaaaattttataacctaggtgataacatagggattggacatgagaca
 25 aataaagtactggatactatatttgaccgggatgatatggaaatgattgcaggtaatcat
 gatgaagctattatgtcactcgtcaatggaacaccttatcctgaagatttaaaaggga
 ttttatgagcatcatcaatggatagaaggacatttagatgagtcctattacgatgaaatt
 aatcaattgcctagatatattgaaatgaccataaaaagggaagattttattttattcat
 tatgaaattgaaaaatgataaaatgtcagctcctattgatgaacaacctttgcacctatt
 30 acaaaagatgacgaacaagctattttctgaattatttaaaagacaaagaagccgatttaatt
 tttatttgacataaccataggttgcatatgtttgatgataaatcaacagtatattttaatt
 cctggatcagtaggtttgaacaatggttcaaacccttatatggtattattaccgttaatt
 gaaaaaggaatttcagtagagagagtgaatttagcgtacaataatgaagaatttttagca
 ggatttgaaagaaaagcaagtagcagctagagaatttatatttaagaatttcatttaa
 35

Sequence 236
 MLEYANKITLKGDNIMKFAVITDIHGDFDALQTVLDDIDSRDDIEKIYNLGDNIGIGHET
 NKVLDTIFDRDDMEMIAGNHDEAIMSLVNGTPYPEDLKGKFYEHQWIEGHLDESYYDEI
 NQLPRYIEMTIKGGKILFIHYEIENDKMSAPIDEQPFAPITKDDEQAISELFDKDEADLI
 40 LFGHNHRLHMFDDKSTVYFNPGSVGLNNGSNTVYGIITVNEKGISVERVKLAYNNEEFLA
 GFEEKQVPAREFIFKNFI*

Sequence 237
 45 Contig_0460_pos_5997_5641,
 putative peptide of unknown function
 atgtttttaataactttattgcctatttttcaatatcaagcttctgcacatgcgacttta
 gaaaaatcaaccacacaacagcaagggttatttaagacaaaaccagaagcaatcaagtta
 gagtttaataatgaacctgtgaacaccaaataactcgagtgtagcttatttgatgataaagg
 50 aaaaagattaaagaccttaaccaataaacaactggatggtctcagacagttgtattttca
 tctgagcaaattgttaatggcacgaataactattgaatggcatagcgtatctgcggatgga
 catgaagtcggagatacggtttgaattttcagttggaaaagtgaggctaaagatgtag

Sequence 238
 55 MFLITLLPIFYQASAHATLEKSTPQQGVKIDKPEAIKLEFNEPVNTKYSSVTLFDDKG
 KKIIDLKPITTGWSQTVVFSSEQIVNGTNTIEWHTVSADGHEVGDTFEFSVGKVRLLKM*

Sequence 239
 Contig_0460_pos_5440_4823,
 putative peptide of unknown function

gtgggtttatatgatgacactcacatctgatatattagaagatattctatcatttaaatta
 gaagtgataatgcaatttccgtatatattaagctctatttcactaatcattttgtttata
 cttttcatttttaaaagatatggaaaaaatatggtagctggtcatttcaatagttatgatt
 5 gctgtgataagtatgtctggacacgtgtgggtcacaacaagtgccattatgggtcaattatc
 ataagaacaattcatcttatagggttaacggttatggtaggttcaactcgttatctcatt
 tgttatgctattaaagtgaaaattaatcagttgacgagtgaagacgtatgcttttaaaa
 gttaatatcattgctgtgattatgctcggtttttacagggttttaattggctattgatgaa
 acgaatactttaacactttggaataatgtgagcgcttgggtctatttatcttgcataaaa
 atcgcaggaattattgctatgatgctattaggtttctatcaaacgatgcgtgctttgaga
 10 caacgacaacaggtccatcggttttgactgatgactgaattgttaattgggtatgatatta
 attttgaggtatcatga

15 Sequence 240
 VVYMMTLTSDILEILSFKLEVIMQFPYILSSISLIILFILFILKDMEKIYWYWLISIVMI
 AVISMSGHVWSQQVPLWSIIIRTIHLIGLTLWLGLSVLYLICYAIKVKINQLTSVRRMLLK
 VNIIAVIMLVFTGILMAIDETNTLTWNNVSAWSIYLVIKIAGIIAMMLLGFYQTMRLR
 QRQQVHRFALMTELLIGMILILQVS*

20 Sequence 241
 Contig_0460_pos_3947_3564,
 putative peptide of unknown function
 atgggtcataggtttattaagtggcttttactacagagaattaactaaagcgcatgacttt
 25 gtgggtgacacgcaattgtcttttagtgcatacacatacacttatcttaggcatgtttatg
 tttttactcttattaccacttgaaaaagtatttaaaattaagtagttattacttatttaaat
 tggttcttttttcgtgtatcatttaggtgtgttaatcacgatttcaatgatgacagttaaa
 ggtacattccaagttattggtaaaaaattttcacccgaaatgtttgcgggatttgcaggc
 ataggtcatacaggtatgcttgcaggtttactgttactgtttttcttattaagacaggct
 30 attcttacagaacccaaaaaataa

Sequence 242
 MVIGLLSGFYRELTKAHDFVGDQTQLSLVHTHTLILGMFMFLLLPLEKVFKLSSYYLFN
 WFFVYHLGLVITISMVTVKGTQVIGKKFSPMFAGFAGIGHTGMLAGLLLLFFLLRQA
 35 ILTEPKK*

Sequence 243
 Contig_0460_pos_2387_882,
 is similar to (with p-value 4.0e-67)
 40 >sp:sp|P35164|RESE_BACSU SENSOR PROTEIN RESE (EC 2.7.3.-). >
 pir:pir|S45560|S45560 hypothetical protein X18 - Bacillus su
 btillis >gp:gp|L09228|BACDIA_27 Bacillus subtilis spoVA to se
 rA region. NID: g410114. >gp:gp|Z99116|BSUB0013_23 Bacillus
 subtilis complete genome (section 13 of 21): from 2395261 to
 45 2613730. NID: g2634723.
 atgagcgagaaaccagactcatcgatataaagacactaaaaacaaatgtttaatgaaata
 aaaaagagtactaaatttaaaaaagtgtttaagaaggtagtatgaaactcaaaatatt
 acaataaaaaataaaaggtaattctcaatcttatcttttgctaggataaccaatgaaagct
 caaaaagggtgctcaaaagtcattatagtggtgtttttatatataaagattgaaatctatc
 50 gaagatacaataatgctattacaattattatattaataactgctattatatttactata
 gcaagcacaatatttgcattctttttatctaataagaataacgaaacccttacgtcaatta
 aaaacacaagcacaaaaagtttctgaaggggattatagtcaaaatttcaactgtcgctact
 aaagatgaaatagggtatttatcacgtgcatttaacaacatgaacgtagaattcaagaa
 catatcaaagcaatttcttcatttaagaatataagagatacattattaaactctatggta
 55 gaaggcggttctaggcatttaataatcaacgtgaaatcatattgtcgaacaagatgggtgat
 gatattatgcgtcacattgatgatttttcaaaagaatctattgaacagcaaattgaagca
 acatttgaatcacacagaatgagttatttagaattagaaattaatacaaggtactatgta
 tttatctccagttatatagatagaattcaacaaatggtagaagtggtattgtcatggtc
 atccgtgatatgacaaatgaacataatcttgatcaaatgaaaaagattttatagcaaat

gtatcacatgaattacgtacgccaatctctttattacagggttacactgagtcacatagta
gacggtatagttaccgaaccagatgaaatacgtgactcattagcaatcgtttagatgaa
tctaagcgacttaatcgtttagtcaatgaattactaaatgtagctcgtatggatgctgaa
5 ggattatcagttgagaaggaattacaacctattcaacaccttcttgataaaatggagctt
aaatatcgcatgcaaagtgaagaattaggtttaacaatgacgtttgattctaataatgac
gaacaattatggaactatgatatggatagaatggaccaagtgttaactaatttaattgat
aacgcaacaagatatacacaagctgggtgattctataaagatttctattgatgaagattca
gatttcaatattttaacaataactgatacaggcactgggtatagcaccggaacatctgaaa
10 caagtatttgaccgtttttataaagtggacgctgctcgaaaaagaggtaagcaaggcacc
ggattaggacttttcatattgtaaaatgattattgaagaacacgggggacgtattgatgtt
gagagcgaattaggcaaaaggtacttcatattattattagactacctaataatcaaaacaaatt
agtttag

Sequence 244

15 MSEKPDSSYKDTKKQMFNEIKKSTKFKKVFKEGEYETQNITIKNKGNSQSYLLLGYPMKA
QKGAQSHYSGVFIYKDLKSIEDTNNAITIIILITAIIFTIASTIFAFFLSNRITKPLRQL
KTQAKVSEGDYSQISTVATKDEIGDLSRAFNMMNVEIQEHIKAISSSKNIRDILLNSMV
EGVLGINNQREIILSNKMADDIMRHIDDFSKESIEQQIEATFESQQNEYLELEINTRYIV
FISSYIDRIQTNGRSGIVMVIRDMTNEHNLDMKKDFIANVSHELRTPIISLLQGYTESIV
20 DGIVTEPDEIRDSLAIVLDESKRLNRLVNELLNVARMDAEGLSVEKELQPIQHLLDKMES
KYRMQSEELGLTMTFDSNNDEQLWNYDMDRMDQVLTNLIDNATRYTQAGDSIKISIDEDS
DFNILTITDTGTGIAPEHLKQVDFRYKVDAARKRGKQGTGLGLFICKMIIIEHGGRIDV
ESELGKGTSTFIIRLPKSKQIS*

25 Sequence 245

Contig_0460_pos_0_368,
is similar to (with p-value 3.0e-18)
>sp:sp|P50726|YPAA_BACSU HYPOTHETICAL 20.5 KD PROTEIN IN SER
A-FER INTERGENIC REGION. >gp:gp|Z99116|BSUB0013_17 Bacillus
30 subtilis complete genome (section 13 of 21): from 2395261 to
2613730. NID: g2634723. >gp:gp|L47648|BACSERA_2 Bacillus su
btilis phosphoglycerate dehydrogenase (serA), ypaA, ferredox
in (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutamat
e dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex lytic
35 enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate kin
ase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent glyc
erol-3-phosphate dehydrogenase (glyc), yphE and yphF genes,
complete cds. NID: g1146195. >gp:gp|L47648|BACSERA_2 Bacillu
s subtilis phosphoglycerate dehydrogenase (serA), ypaA, ferr
40 edoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glut
amate dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex l
ytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate
kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent
glycerol-3-phosphate dehydrogenase (glyc), yphE and yphF gen
45 es, complete cds. NID: g1146195.
atgggagaagatggaggtttttgtgttcaaatttcctccttttctaagtaagatgaat
ggaaggagaaaattatatatgcaacaaaacggtttgattacaattagatgttaagt
gcggtagcgtttgtgttaactttcatcaagtttcattgccatttataaccaccgtatcta
actctcgatttttagtgatgtaccgacgttatttagcaacattcctcttaagtcctattgct
50 gggattatcgttgactcatcaaaaatattttaaattttctattcaatataggggaccc
gttgaccagtagctaacttttttagcaggcgtcagctttttgctatcatcactatgtt
tatTCTAT

Sequence 246

55 MGEDGGFLLFKFPFLSKMNGRRKLYMQQNKRLLITISMLSAVAFVLTFIKFPLPFIPPYL
TLDFSDVPTLLATFLLSPIAGIIVALIKNILNLFNIGDPVGPVANFLAGVSFLLSSYYV
YSX

Sequence 247

Contig_0461_pos_160_477,
putative peptide of unknown function
atggcaaatgcggtgacggttttaagtgggtgacagttatttcatgtcagaacttaagcaa
ctgggtacaccagaggtatgtccatcccggtcctctcgtactaaggacagctcctctcaaa
5 tttcctacgcccacgacggatagggaccgaactgtctcagcagcttctgaacccagctcg
cgtaccgctttaatgggcgaacagcccaacccttgggaccgactacagccccaggtatgagc
atgagccgacatcgaggtgcccacacccctcccgcgtatgtgaactcttgggggagataagc
ctggtatccccgggtag

10

Sequence 248
MANADDVLSGDSYFMSELKQLVHQRYVHPGPLVLRTPALKFPTPTTDRDRTVSRSEPPSS
RTALMGEQPNPWDRLPQDAMSRHRGAKPPRRCELLGEISLLSPG*

15

Sequence 249
Contig_0461_pos_4273_4671,
putative peptide of unknown function
gtgcatagtacttacacatttgttcttccctaataacagagttttacgatccgaagacc
20 ttcactcactcacgcggtgtgctccgtcaggctttcgccattgcggaagattccctact
gctgcctcccgtaggagtctggaccgtgtctcagttccagtggtggccgatcaccctctca
ggtcggctacgcatcaattgtggtgcgttattatttttcatgatttttaggtcaggtcaat
tatttttatggtattattatggcttctagcatgataggtgcgttgttaggtgctcaa
tttgctttgaaaaaagggttaggatatgtaaaagctttatttttagtggttactgcaata
25 ttaattataaaaaatctctacgattttattgtgcagtaa

Sequence 250
VHSYLHICSSLITEFYDPKTFITHAALLRQAFHCGRFPTAASRRSLDRVSVPVWPITLS
GRLRINC GALLFFMILGQVNYFYGIIMASSMMIGALLGAQFALKKGVGKALFLVVTAI
30 LIIKNLYDFIVQ*

Sequence 251
Contig_0461_pos_4914_5408,
is similar to (with p-value 1.0e-41)
35 >gp:gp|L11577|STRSCAA_5 Streptococcus gordonii coaggregation
mediating adhesin (scaA), ATP binding protein, hydrophobic
membrane protein, complete cds, and zinc metalloprotease gen
e, partial cds. NID: g310629.
atgactcaaattacatttaaaaaataatcccattaaattatcaggttctgaagtgaatgaa
40 ggtgatatcgaccaaatttcacagtgcttgataatagtttgaaatcaaattacttttagat
gattataaaaaacaaaagaaatttaattagtggtataccatctattgatacaggagtatgt
gatagtgcaaaactcgaaagtttaataagaagcttcagcagaagatggtgtagttttaacg
atatcagtagattttacctttcgcccaaaaaagatggtgtgcatcaagcggattagataat
gtaattactttaagtgatcataaagatttatcttttggtcgaaattatggacttgtgatg
45 gatgaattacgcttacttgacggttcgggtatttgtgttaaacgaaaacaataaagtagta
tataaggaaattgtcagcgaaggtacgaattaccctgattttgaagctgcattaaaagct
tacagaaatatttag

Sequence 252
50 MTQITFKNNPIKLSGSEVNEGDIAPNFTVLDNSLNQITLDDYKNKKKLISVIPSIDTGVC
DSQTRKFNEESAEDGVVLTISVDLPFAQKRWCASSGLDNVITLSDHKDLSFGRNYGLVM
DELRLRLARSFVLNENNKVVYKEIVSEGTNYPDFEALKAYRNI*

Sequence 253
55 Contig_0461_pos_5504_6484,
is similar to (with p-value 7.0e-52)
>sp:sp|P37876|YTXK_BACSU HYPOTHETICAL 37.4 KD PROTEIN IN ACK
A-SSPA INTERGENIC REGION. >gp:gp|AF008220|AF008220_144 Bacil
lus subtilis rrnB-dnaB genomic region. NID: g2293135. >gp:gp

|Z99119|BSUB0016_21 *Bacillus subtilis* complete genome (section 16 of 21): from 2997771 to 3213410. NID: g2635411.
 atggacacttttttaaatagaaagggatattttatgtctgaagaaaatactattatggaa
 cgtctatttcataaattagatgataaaagctaaaacgttaacaaaagaaaatggacagagt
 5 tttatcgaaaatttagggtagctatggaagatatttatacaaaccaagagaactttta
 gaacaagcaacgcttcaagatagaaggaaagcttttcaatttgcataatttaagttatta
 caagaagaaaatattcaagctaattcatcagatcacgcctgactctataggactcattctc
 ggttttcttggttcaacgcttttttagaacataaaaaaggaaatgcacattgtagatattgca
 agtggggcgaggtcatctaagtgcagctgtgaaagaagtactttctgataaaacaattatg
 10 catcatctgatagaggttagatccagtgctatcacgtgtaagtgtgcatttggctaatttt
 ttagagataaccgtttgacgtttatcctcaagatgagattatgccattaccattggaagag
 gctgatgtcgtgattggagatttcccaataggatactatccttttagatgaacgtagtaga
 gaaatgaagttaggctttgaagagggaacaggtattcccatcatctgttaatagaacaa
 tctattaatgcgctaaaaggggcaggttatgcatttttagttgttcttagtcatctcctt
 15 taagatgataaaagtgaacagttggaaaatttcattgctacagagactgagatgcaagca
 tttttaaatttacctaaaacattatttaaaaaatgaaaagcacgtaaatctatattgatt
 ttacaaaagaaaaatcaggcgaaactcgaccagttgaagtcctattagccaatatccct
 gattttaaaaatcctcaacaatttcaagggtttcatttctgaattgaatcagtggaatagtc
 acaaatcatacaaaaaaatag

20 Sequence 254
 MDTFLNRKGYFMSEENTIMERLFHKLDDKAKTLNKENQSFIEENLGLAMEDIYTNQRELL
 EQATLQDRRKAFQFAYLSLLQEENIQANHQITPDSIGLILGFLVQRFLEHKKEMHIVIDIA
 SGAGHLSAAVKEVLSDKTIMHHLIEVDPVLSRVSVHLANFLEIPFDVYPQDAIMPLPLEE
 25 ADVVIGDFPIGYPLDERSREMKLGFEEGHSYSHLLIEQSINALKGAGYAFLVVPShLL
 EDDKVKQLENFIATETEMQAFNLNPKTLFKNEKARKSILILQKKKSGETRPVEVLLANIP
 DFKNPQQFQGFISELNQWIVTNHTKK*

Sequence 255
 30 Contig_0461_pos_3582_3238,
 putative peptide of unknown function
 gtgacaaaccggaggaaggtggggatgacgtcaaattcatcatgcccttatgatttgggc
 tacacacgtgctacaattggacaatacaaaaggcgagcgaaaccgaggtcaagcaaattcc
 cataaagttgttctcagttcggattgtagtctgcaactcgactatatgaagctggaatcg
 35 ctagtaattcgtagatcagcatgctacggtgaatacgttcccgggtcttgtacacaccgcc
 cgtcacaccagagagtttgaacaccgaagccggtggagtaaccatttggagctagcc
 gtcgaaggtgggacaaatgattgggggtgaagtcgtaacaaggttag

Sequence 256
 40 VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSDDCSLQLDYMKLES
 LVIVDQHATVNTFPLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVVTR*

Sequence 257
 Contig_0462_pos_27_440,
 45 is similar to (with p-value 5.0e-18)
 >gp:gp|AF012906|AF012906_6 *Bacillus subtilis* yojP gene, part
 ial cds; yojQ/S, yojR, yojT, yojU, yojV, yojW, yojX, yojY, y
 ojZ, and yokA genes, complete cds. NID: g2522404. >gp:gp|Z99
 114|BSUB0011_163 *Bacillus subtilis* complete genome (section
 50 11 of 21): from 2000171 to 2207900. NID: g2634230. >gp:gp|AF
 020713|AF020713_166 Bacteriophage SPBc2 complete genome. NID
 : g3025478.
 atgataggtacttatcaaagtgataaaaactttgaaatgatgaagacttttaagcattgg
 attcagactaatcattattggaaatatgttgagaaatcgggtgtgttaggtatagcatta
 55 gataatcctctccacgttcaaagtaataatgtagatatgacgttgttttgagaatagat
 gaaacagtaaatgatcagacaatatctaaaagagattttacaggtggcatatatgtctgtg
 tttaaagttagtcatacaaaaataaatatagagaagttcttttagcaatttagaaaatatt
 ttaaatgaaagtcatttgcgtatgagaaatgaaccaattatagagagatacattgaagaa
 gagggacagataaaagtgtgtgaaatgttagtgcctatctatgaagtaaatata

Sequence 258

MIGTYQSDKNFEMMKTFKHWIQTNHYWKYVEKYGVLGIALDNPLHVQSNQCRYDVVLRID
ETVNDQTISKRDFTGGIYAVFKVSHTKINIEKFFSNLENILNESHLMRNEPIIERIYEE
5 EGTDKVCEMLVPIYEVN*

Sequence 259

10 Contig_0462_pos_1824_3440,
is similar to (with p-value 3.0e-53)
>sp:sp|P45082|CYDD_HAEIN TRANSPORT ATP-BINDING PROTEIN CYDD.
>pir:pir|F64186|F64186 transport ATP-binding protein (cydD)
homolog - Haemophilus influenzae (strain Rd KW20) >gp:gp|U3
15 2795|U32795_6 Haemophilus influenzae Rd section 110 of 163 o
f the complete genome. NID: gl574708.
atggttttaagttataaattataccccacactcatgttaatcatgagcggttttttatct
tttacggctcggtgcgcaaaacatttcaatttcacactttttaatcacttactgtattat
caacaacaatctttattattattgttatcagttattttatctctcttattttaagagca
acatttaatatgctgattcaatttttaggagatcatttggcatttaagtaaaacatatg
20 cttagagaacaagtgtatttgaaaaaagtgccgttcaattggtgaagaaataaatatt
ttaactgaaagtattgatggtatcggtccggtctttcagagttatttacctcaagtcttt
aatcaatgttgattcccatcggtattattattaccatgtgtttgttcatttacctact
gctattattatgatagttaccgcaccttttattccattgtttatgttatttttgactt
25 aaaacaagagatgagtc aaaggatcaaattgacatatttaaaaccagtttagtcaacgtttt
ttaaatacagctaaaggctcttattacatttaaaacttttaaatcaaacgaaacaatctgag
caacaactttataaagacagtcacacgttttagagatttaacaatgcgtattttgaaaagt
gcctttttatcaggacttatgcttgagttcataagtatgttagggattggattggtcgca
ttggaagcggttttaagcttagttgtatttaaccatatcaactttgtgactgcagcgata
30 gcgattatttttagctcctgaattttataatgcgattaaagatttaggtcaagcatttcat
acaggtaagcaaaagtgaaggtgctagcgatgtggtgttttcatttttagaatctgaagat
aaagctgattctcctacattaaaagtggatgagcaacagtttgaacaagttttaattaa
catgttgattttcaatacgctaataagtaatacatatggctttgaaaaacatttcttttcg
gtaataaaaggagaaaaaggtcgctattgtgggaccgagtggtgcagggaaatccacttta
35 gctaagttgcttagtcaatcagtaaacacccacacatggaacactttcatttaaccaagca
tcattaaatatcggttttctaagtcagcgccacatatatttgcagattctatcaaaaat
aatattgcaatgtatgatgatgagatatgtgatgagcaagtgattcaagtgccttgatgaa
gtgggggttaaaagagaaagtactttcattaaaatatggtatctatacttctattggtgaa
ggtggggaaatgttatcaggtggacaaatgagacgtattgagtttaagtcgtttattatta
40 ttgaaaccagatatgttaatttttgatgaaccagcgataggattagattgaaactgaa
aaggtcatacaacaagtattagagcatcatttttctacaacgacagtggttattattgca
caccgtgattcaaccattcgaagttcagcacggcgatatatatcgaaagtggtcatctt
ataaaagatgattcgataatttctgttacgcgtagtgggtgaagatagatcaatga

45 Sequence 260

MVLSYKLYPTLMLIMSVFLSFTVVAQNISISHFLNHLLEYQQSLLLLLSVIFISLILRA
TFNMLIQFLGDHLAFKVKHMLREQVILKKS VRSIGEEINILTESIDGIGPFFQSYLPQVF
KSM LIPIVIIITMCFVHLPTAIIMIVTAPFIPLFYVIFGLKTRDESKDQMTYLNQFSQRF
LNTAKGLITFKLLNQTKQSEQQLYK DSTRFRDLTMRILKSAFLSGLMLEFISMLGIGLVA
50 LEAALSLVFNHINFVTA AIAIILAPEFYNAIKDLGQAFHTGKQSEGASDVVFSFLESED
KADSP TLKVDEQQFEQVLIKHVD FQYANSNHMA LKNISFSVNKGEKVAIVGPSGAGKSTL
AKLLSQSVTPHGTLSFNQASLNIGFLSQRPHIFADSIKNNIAMYDDEICDEQVIQVLDE
VGLKEKVL SLKYGIYTSIGEGGEMLSGGQMRRIELSRLLLLKPDIVIFDEPAIGLDIETE
KVIQQVLEHHFSTTTVFIIAHRDSTIRSSARRIYIESGH LKDDSIISVTRSEVKIDQ*
55

Sequence 261

Contig_0462_pos_1513_728,
is similar to (with p-value 6.0e-63)
>gp:gp|U87792|BSU87792_1 Bacillus subtilis tRNA-Ala, phospho

tidylglycerophosphate synthase (pgsA) and ClnA (cinA) genes, complete cds, and RecA (recA) gene, partial cds. NID: gl842434.

5 atgatacttgtcgatgatatgtgggttaaagtcactaatcttctcggttctcaatcagca
ttcacatttaaagttgtttatacagtttaggttcagtatcttgcggccgctgggtatttaga
gaacgcttcttagaaattttacatattggccaacataaacctgaaccttccacttcggga
gacagcgttcaaaaccacgacgtctgaatttaatacatgtattagtaggtatggccca
gcagggtatttaggtattttatttgatgatttaattgaaaaatacttatttagtgacca
acagtcttaattgggtttatttataggtgccatttatatgattatagctgataagtattct
10 aaaactgttcagcatcctcaaacagtagatcaaattaattttccaagcatttgcatt
ggatctctcaagcaatagctatgtggcctggatttagtagatccggttcaacgatttca
acagggtgttcttatgaaattgaatcataaagctgcattctgatttcaacttttattatgtcg
gtaccaattatgttagctgcaagtggtattatcttactaaaacattatgagtatattcat
ttagcacacataccattctacatttttaggatttttagcggcatttattgttggttaatt
15 gcaattaaaacattcttacacttaattataaagtttaagttagtagcttttgcatttat
agaattgtcttagttatttttatagcaatcctatacttcggatttcggtattggcaaagga
atttaa

Sequence 262

20 MILVDDMWLKSTNFLGSQSAFTFKVVIQLGSVFAAAWVFRERFLEILHIGQHKPEPSTSG
DRRSKPRRLNLIHVLVGMVPAGILGFLFDDDLIEKYLFSVPTVLIGLFIGAIYMIADKYS
KTVQHPQTVDQINYFQAFVIGISQAIAMWPGFSRSGSTISTGVLMLNKAASDFTFIMS
VPIMLAASGLSLKHIEYIHLAHPFYILGFLAAFIVGLIAIKTFLHLINKVKLVFFAIY
RIVLVIFIAILYFGFGIGKGI*

Sequence 263

Contig_0463_pos_4479_6836,
is similar to (with p-value 0.0e+00)
>gp:gp|M86227|STARECF_3 Staphylococcus aureus DNA gyrase B s
30 ubunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit
(gyrA) gene, complete cds. NID: gl53083.
gtgtcctgtactatggatgggtgacggtgcagctgcaatgcgttataccgaagcacgtatg
actaaaataacattagaacttttacgtgatattaacaaagacacaattgattttattgac
aactatgatggtaaatgaaagagagccgctcagcttacctgcacgtttccctaacttacta
35 gtaaatgggtgcggcaggaattgccgtaggtatggctacaaatattcctccccacaattta
actgaagttattgatgggtgtgctcagtttaagtaagaatccagacatcacaattaatgag
ctgatggaagacatacaaggtcctgattttcctacagctgggttagtactagggaaaagt
ggatttcgtcgagcttatgaaacaggtcgtgggtcaattcaaatgcgttctcgtgctgaa
atagaagaacgtgggtgggtggccgtcaacgtattgtcgtaacggaaatacctttccaagtc
40 aataaagcgcgtatgattgaaaaaatcgcagagtttagtttagagataagaaaatcgacggt
attacagatttacgtgatgaaacaagtttgcgtacaggtgtaagagttagttattgatgta
cgtaagatgcaaatgcgagtggtatttttaataatttatataaaacaaacgccattacaa
acatcatttgggtgtgaatatgattgcttttagtgaatggtagacctaaactaatcaattta
aaagaagcacttatccattacttagaacacccaaaaaacagtgggttagacgacgtactgaa
45 tataatcttaaaaaagcaagagaccgtgcccatattctagaaggtttacgaatagcacta
gatcatattgatgaaattatcacaacaattcgtgaatcggacactgataaaattgcgatg
gcaagtttacaagagcggttttaactaactgaacgtcaagctcaagcaatttttagatatg
cgttttaagacgttttaactggattagaaagagataaaatagaatctgagtataatgaactt
ctagaatatattaaagagttagaagagatttttagctgatgaagaagtactattacaatta
50 gttcgtgatgaattgactgaaattaaagaacgtttcggcgatgaacgtcgactgaaatt
caattaggtgggtctagaagatcttgaagatgaagacttaatccctgaagaacaaattggt
attacattaagtataataactatattaacgtttaccagtatctacatatcgttctcaa
aatcggtgggtcgtggcatacaaggtatgaacacgttggatgaggacttcggttagtcaa
ttggttaacaatgagtacacatgattatgttctgttctttacgaataaaggtcgtgtatat
55 aaactcaaaggttatgaagttcctgagttgtcacgtcaatccaaaggcatacctattatt
aatcgcatgtaactcgaaaaatgacgaaacaataagtagcatgattgcagttaaagacctt
gaaagtgaagaagattatctcgattttgcgacaaaacaaggtatcgttaaacgttcatca
ttaagtaacttctcccgatttaacaaaaacggtaaaattgcaattaaactttaagaagat
gatgaattaattgcagtagcttaacaacaggtaatgaagatattcttattggaactgca

catgcatcattaattagattctctgaatctacattacgcccattaggccgtacagcagca
 ggtgtgaaaggtatttctctacgtgaaggggatactgtcgtaggtcttgatggtgcagat
 tcagaaagtgaagatgaagtattagtagttactgaaaatggttacggtaaacgtacacct
 gttagcgaatatcgtttatcaaactcgtggtggttaaaggaatcaaaactgcgacaattacc
 5 gagcgtaatggtaacatcgtttgtatcacaactgtaaccggtgaagaggatttaattggtt
 gtaactaacgctggtgttattattcgtcttgacgttcatgatatttctcaaaatggacgt
 gcagcacaaggtgtacgccttatgaaactcggagatggtcaatttgtttctactgttgct
 aaagttaaacgaagaagacgataatgaggaaaatgcagatgaagcgcaacaatctactact
 actgaaacagcagatgtagaagaggtagtctgatgatcagacaccaggcaatgcgattcat
 10 acagaaggtgatgcagaaatggaatctgtagagtttctgaaaatgatgatcgtattgat
 atcagacaagattttatggatagagtgaatgaagatatcgagagtgttccagataatgaa
 gaagatagtgtgaataa

Sequence 264
 15 VSCTMDGDGAAAMRYTEARMKITLELLRDINKDTIDFIDNYDGNEREPSVLPARFPNLL
 VNGAAGIAVGMATNIPPHNLTEVIDGVLSLSKNPDITINELMEDIQGPDPFTAGLVLGKS
 GIRRAYETGRGSIQMSRAEIEERGGGRQIVVTEIPFQVKNKARMIEKIAELVRDKKIDG
 ITDLRDETSRLTGVVVVIDVRKDNASVILNNLYKQTPLOTSFGVNMIALVNGRPKLINL
 KEALIHYLEHQKTVRRRTEYNLKKARDRAHILEGLRIALDHIDEIITTIRESDTDKIAM
 20 ASLQERFKLTERQAQAILDMRLRLTLGLERDKIESEYNELLEYIKELEEILADEEVLLQL
 VRDELTEIKERFGDERRTEIQLGGLEDEDEDLIPPEQIVITLSHNNYIKRLPVSTYRSQ
 NRGGRGIQGMNTLDEDFVSQVLVTMSTHDYVLFFTNKGRVYKLGKYEVPESRQSKGPII
 NAELENDDETISTMIAVKDLESEEDYLVFATKQGIVKRSSLNFSRINKNGKIAINFKED
 DELIAVRLTTGNEDILIGTAHASLIRFSESTLRPLGRTAAGVKGISLREGDTVVGLDVAD
 25 SESEDEVLVVTENGYGKRTPVSEYRLSNRGGKGIKTATITERNGNIVCITVTGEEDLMV
 VTNAGVIIRLDVHDISQNGRAAQGVRLMKLGDGQFVSTVAKVNEEDDNEENADEAQOSTT
 TETADVEEVVDDQTPGNAIHTEGDAEMESVEFPENDRIDIRQDFMDRVNEDIESASDNE
 EDSDE*

30 Sequence 265
 Contig_0463_pos_8889_10175,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P95689|SYS_STAAU SERYL-TRNA SYNTHETASE (EC 6.1.1.11)
 (SERINE--TRNA LIGASE) (SERRS). >gp:gp|Y09924|SASERS_1 S.aure
 35 us serS gene. NID: g1835217.
 atgttagacattcgtttatttagaaatgaacctgagaaagtgaagagcaaaattgaatta
 agaggcgacgatctaaagttgtcgaccaagttttagaattagatgaacaacgccgtgaa
 ttaatcagtaaaactgaagagatgaaggcgaaaagaaataaagtgaagcgaagaaatagct
 caaaagaaacgtaataaagaagacgctgatgatgtcattgctgagatgcgtcatttaggt
 40 gatgaaattaaagatatcgataatcaacttaataatgaagttagataataaaattagagatc
 ttaattcgtattcttaacttaattaatgaagtgtacctcaaggtgattctgatgaagaa
 aacgttgaagttaaaaaatggggtacgccacgtgattttgaatttgaacctaaagcgac
 tgggatttagttgaagaattaaaaatggctgactttgaacgtgctgctaaagtatctggt
 gctcgtttcgtatacttaactaaagatggcgacattactgaacgtgctttaatgaattac
 45 atgttgacaaaacatacaacgcaacatggttatactgaaatgatgacacctcaattagtg
 aatgctgatacgtatgttggaaacaggtcaattacctaatttgaagaagatttatttaa
 gttgaaaaagaaggcttatatacgaattccaactgcagaagtacctttaacaaacttctat
 agagatgaaattattcaaccaggtgtactacctgaattattacagctcaaactgcatgt
 ttccgttagtgaaagcaggtacagctggttagagatactagagggttaattcgtttacatcaa
 50 tttgataaagttgaaatggttcgtattgtacaacctgaagattcttgggatgctttagaa
 gaaatgacacaaaaatgctgaagctattcttgaagaattaggtttaccataaccgtcgtgtt
 atcttatgtactggcgatattggtttcagtgctagtataacatatgatttagaagtttg
 ttaccaagttacaatgattataaagaaatcagttcttgccttaactgtactgatttccaa
 gcacgtcgcgcaaatatcagattcaaacgtgatgctgcttctaaaccagaatttagtacac
 55 acattaaatggtagtggtttagcagtaggtcgtacatttgcagccatcgttgaaaactat
 caaaacgaagatggtacattacaatttctgaagcatttagtaccatttatgggtggcaaa
 actaaaattgaaaaaccaatcaataa

Sequence 266

MLDIRLFRNEPEKVKSKIELRGDDPKVVDQVLELDEQRRELISKTEEMKAKRNKVSEEIA
 QKKRNKEDADDVIAEMRHLGDEIKDIDNQLNEVDNKIRDILIRIPNLINEDVPQGDSDEE
 5 NVEVKKWGTPRDFEFEPKAHWDLVEELKMADFERAAKVSGARFVYLT KD GALLER ALMNY
 MLTKHTTQHGYTEMMPQLVNADTMFGTGQLPKFEEDLFKVEKEGLYTIPTAEVPLTNFY
 RDEIIQPGVLPFLFTAQACFRSEAGSAGRDTRGLIRLHQFDKVMVRIVQPEDSWDALE
 EMTQNAEAILLEELGLPYRRVILCTGDIGFSASKTYDLEVWLP SYNDYKEISSCSNCTDFQ
 10 ARRANIRFKRDAASKPELVHTLNGSGLAVGRTFAAIVENYQNE DGTLT IPEALV PFMGGK
 TKIEKPIK*

Sequence 267

Contig_0463_pos_12074_12766,
 putative peptide of unknown function
 15 atgactcaccttacgttttaacaagggtgtgaaagagtgtattcccacgttacttggttat
 gcagggtgtaggactatcgtttggaattgtggcagtcctcccaaaatttcagtgttttagaa
 attattttattgtgtctgattatttatgctggcgagctcaatttattattgtacatta
 gtgattgcaggcacccttatttctgcaattgtgcttacaatacttatcgtttaactctcga
 atgttcttattaagtatgacttttagcacctaattataagcaaatatggattttggaatagg
 20 gtagggcttggaacgtttattaacagatgaaacttttggcgttgctataaacaccatatgtt
 aaaggtgaaaaaatttaacgatcgatggctacacggactaaatattactgcttacttattt
 tggactgtttcctgtgtaatcggtgccattttcgagaggtatatttcaaactcctgatgag
 ctgcgcctagactttgcaattaccgcaatgtttatttttttatgtatatctcaatttgaa
 gggattaagaaatcacgattgagaatatattgtactcattgtatgtgtgattgtgatg
 25 atgcttcttctaagttcaattctaccttcatacctagcaatttttaataagccgcaattgtt
 gctgcattgttaggggtggtgatggacaaatga

Sequence 268

MTHLTFKQGVKECIPTLLGYAGVGLSFGIVAVSQNFSVLEIILLCLIIYAGAAQFIICTL
 30 VIAGTPISAIIVLTILIVNSRMFLLSMTLAPNYKQYGFWRVGLTLLTDETFGVAITPYV
 KGEKINDRWLHGLNITAYLFWTVSCVIGAI FGEYISNPDALGLDFAITAMFIFLCISQFE
 GIKKSRLRIYIVLIVCVIMMLLLSSILPSYLAILIAAIVAALLGVVMDK*

Sequence 269

Contig_0463_pos_13381_14349,
 is similar to (with p-value 2.0e-26)
 >gp:gp|Z98271|MLCB1779_3 Mycobacterium leprae cosmid B1779.
 35 NID: g2326678.
 atgacgaattacacgggttaatacattagaactaggtgagtttaaaactgaatctggtgaa
 40 acgattgatcattttacgtctacgttatgaacatgtaggacttcttggtcaacccttgctc
 gttgttttgccatgcacttactggcaatcatttaacatacggcacggatgcacaacctggc
 tgggtggcgagaaatcattgacgggtggctacattccagttcatgattatcaatttcttaca
 ttcaatgtcattggaagtccatttgggttcgagttctaaattaaatgatgataacttccca
 gaacatttaacattgagagatattgttagagctattgagttaggtatacaagcattagaa
 45 ttaagaaaattaatattctcattggaggtagtttaggtggtatgcaagcgatggaattg
 ctttataatcgtcaattcgaggtggaaaaagcaatcatattagctgctactgataaaacg
 tcctcttatagtcgtgcttttaacgagattgcaagacaagctatacatataggcggtaaa
 gaaggtttaagtattgcacgtcaactcggctttctcacgtatcgatcgctctaaaagttat
 gatcaacggttttacaccagatgaagtagtgagctatcaacaacatcaagggtgataagttc
 50 aaagaatatttctgatttaaatgtttatttaacactgctagacgtcttagatagtcacat
 ttagatagaggaagagatgatgttgatgaagtctttcagtcggttggaacgaaagtacta
 acaatgggttttattgacgatttgctttatcctgatgatcaagtgagagccttaggagaa
 cgtttttaaatatcatcgctcatttcttcgtaccagataatgtgggacatgatgggtttctt
 55 ctaaaatttaagattgggcgcctaatttatatcatttcttaaaattgaaacaattccga
 cgtaaatag

Sequence 270

MTNYTVNTLELGEFKTESGETIDHLRLRYEHVGLPGQPLVVVCHALTGNHLYGTDAQPG
 WWREIIDGGYIPVHDYQFLT FNVI GSPFGSSSKLNDDNFPEHLTLRDI VRAIELGIQALE
 FKKINILIGGSLGGMQAMELLYNRQFEVEKAIILAATDKTSSYSRAFNEIARQAIHIGGK
 EGLSIARQLGFLTYRSSKSYDQRF TPDEVVSYQQHQGDKFKEYFDLNCYLTLDDVLDSSH
 5 LDRGRDDVDEVFQSLETKVLTMGFIDDLLYPDDQVRALGERFKYHRHFVFPD NVGHGDFL
 LNFNDWAPNLYHFLNLKQFRRK*

Sequence 271

Contig_0463_pos_14578_15504,
 10 putative peptide of unknown function
 atgggagtggcgtatgtgttttcaaaaatacaacctaagcaactattttagcaattatt
 tcattatttagtagtcgcttttagtaacacatgtattacctgttctcggttgattttatgt
 ttatttgcacagattcccggtattgttttgtggaatcggtccatacaatcattcggaatt
 agtgcattagtaacagttgtacttacaacattattaggaatacatttgccttaagtatg
 15 atggctcttaattctattattaagtgcgattatcggaacaattacttaaggaaagaacatct
 aaagaacgaattctttatatttcaacagcttcattagtttagttacacttattggatgg
 atgttattacaacatttcgataaaattccgacggcagctgtattaattaagcctttaaaa
 aatgcaatgcatgaggctttcttaaaaagtggaaatcgattcaaaactatagacagattctt
 gaggaaagtttccgacaaatgacggtccaactccctagttttctaattatagttattttc
 20 atttttgtctttaattatctgattattacatttccaattttacgcaaaatttaagtagca
 acacctatttttaaacctttattcgcattggcaaatgagccgtaatttactatggttttat
 cttatagtagtattttgtgtcatgattgcgagtgaaaccaagtacgttccaaagcatcgtg
 ttaaaactttgatgtgtgttatcattagtgatgtacatccaaggattaagtgtcattcac
 ttcttttggtaaaagctaaaagatggccgaactttgcaacaattcttgttatggtagtaggg
 25 acgcttcttacaccggcaacgcataattgttgattacttggggtaattgatttatgtatt
 aatttaagaaaaataataaaaaaatga

Sequence 272

MGVAYVFSKI QPKATILAIISLLVVALVTHVLPVLGLILCLFATIPGIVLWNRSIQSFGI
 30 SALVTVVLTTLLGNTFVLSMMVLILLLSAIIGQLLKERTSKERILYISTASLSLVTLLIGW
 MLLQTFDKIPTAAVLKPLKNAMHEAFLKSGIDSNYRQILEESFRQMTVQLPSFLIIVIF
 IFVLINLIITFPILRKFKVATPIFKPLFAWQMSRNLWFLIVLIVLICVMIASEPSTFQSIV
 LNFDEVLSLVMYIQGLSVIHFFGKAKRWPNFATILVMVVGTLTTPATHIVGLLGVIDLCI
 NLKKIIKK*

Sequence 273

Contig_0463_pos_15515_17500,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P37484|YYBT_BACSU HYPOTHETICAL 74.3 KD PROTEIN IN RPL
 40 I-COTF INTERGENIC REGION. >gp:gp|D26185|BAC180K_10 B. subtil
 is DNA, 180 kilobase region of replication origin. NID: q467
 326. >gp:gp|Z99124|BSUB0021_156 Bacillus subtilis complete g
 enome (section 21 of 21): from 3999281 to 4214814. NID: q263
 6442.
 45 gtgattagaggtggaagaatgaaccgtcaatccactaaaaaagctttgctcataccggtt
 attttaatggtgctcactgctatagcacttgctcgccgtgtggtttatttttaaccaacta
 gtggcaggtattgctacagctatacttattgtgatgattattattagtggcgtgttattg
 agaaaagcatttctaaaaatggataattatgtggatgatttaagtggtcacatctcgga
 agtagtaacaaggcgattaagcacttgctatagggatgattgtggttagatgaagataat
 50 cacattgagtggtgaaccaatttatgacagatcacattgaaacgaatgtgatttctgaa
 aatgtcaatgaagtcttccctaacatattaaaacaactggaaaaagtccaagaagtagaa
 atagaaaacaacaattattactatcatgtacgatattcagaaaacgagcattgtttat
 ttctttgatatgactgaaactgaacgtacaaacgaactatatgaagattcaaaaccgatt
 attgcaacaatatttttagataattacgatgaaatcactcaaaacatgaacgatacaca
 55 cgttctgaaattaaactctatggtgacacgtgtgattagtcggtgggcacaggattacaat
 atttacttcaaacggttacaactcagatcaattttagcttactttaacaaaaaatattg
 gctgaattagaagattctaattttgaaatcttaagccaattaagagaaaagagtggtgggt
 taccgcgcacaactaacattaagtattggtgtaggtgaaggtagtagaaccttattgat
 ttaggtgaattatcacaatctggttagacctcgcttaggtcggtggtgaccaaagt

gcaattaagaatatgaacggcaatgtaagattctatgggtggaagactgaccctatggaa
 aaacgtacgcgtgtacgtgcggtgtgatttcacatgccctcaaagatatcttactgaa
 ggcgataaagttatcgttatgggacataaagcgaccagatttagatgctataggtgcagct
 atcggagtttcgcgctttgcatcaatgaataatttagaggcatttatcggtcttaatgat
 5 tctgatattgatccgacattacgtcgtgttatggacgagattgataagaaaccggaacta
 aaagaacgcttttgaacatcggtatgaggttgggatatgatgacttctaagacgactgtc
 gttgtgtgatacacataaaacctgaaatggcttagatgaaaatgtcttaataaagca
 aaccgcaaaagtagtcattgatcatcatagacgtggcgaaagctttatttcaaatccatta
 cttgtgtatatggaaccttacgctagctcaactgctgagctcgtaacggaattactagaa
 10 tatcaaccaactgaacagagattgactcgtttagaatcaactgtcatgtatgcaggatt
 atagtagatacaagaactttactttaagaacaggttccagaacatttgatgccgcaagt
 tatttacgtgcacatggcgctgatacaatcttaacgcagcatttcttaaaagatgatgtc
 gatacgtatatcaatcggttcagaattgataagaacagttaagatacaagatcaaggtgta
 gccattgcacatgggttcagatgataaaatttatcatcctgtaacggttgcaaacgtgcc
 15 gacgagttgttaagtttagaaggcattgaagcatcttatgtagtagctaaacgtgaagac
 aacctgatcggtatctcagcacgttcattaggttccataaatgttcaattaacaatggaa
 gcgttaggtggcggtggccatctgacaaatgctgacacaaaataaaaggtgcgacaata
 gatgaagcaatagaacaattacaacaagcaattacagaacaaatgagtaggagtgaaagac
 gcatga

Sequence 274
 VIRGRMNQRSTKKALLIPFILMVLTAIALVAVWFIFNQLVAGIATAILIVMIIISGVLL
 RKAFILKMDNYVDDLSGHISASSNKAIKHLPIGMIVLDEDNHIWNNQFMTDHIETNVI
 NVNEVFNNILKQLEKVEVEIENNNYYYHVRYSENEHCLYFFDMTETERTNELYEDSKPI
 25 IATIFLDNYDEITQNMNDTQRSEINSMVTRVISRWAQDYNIFYKRYNSDQFVAYFNQKIL
 AELEDSNFEILSQLREKSVGYRAQLTSLIGVGEGTENLIDLGELSQSGLDLALGRGGDQV
 AIKNMNGNVRFYGGKTDPMKRTVRARVISHALKDILTEGDKVIVMGHKRPDLDAIGAA
 IGVSREASMNNLEAFIVLNDSDIDPTLRRVMDEIDKKPELKERFVTSDEAWDMMTSKTTV
 VVVDTHKPEMVLDENVLNKANRQVVIDHHRGESFISNPLLVMEPYASSTAEVLVTELE
 30 YQPTQRLRLESTVMYAGIIVDTRNFTLRTGSRTFDAASYLRAHGADTILTQHFLKDDV
 DTYNRSSELIRTVKIQDQGVIAHGSDDKIYHPVTVAQAADELLSLEGIEASYVVAKRED
 NLIGISARSLSINVQLTMEALGGGGHLTNAATQIKGATIDEAIEQLQQAITEQMSRSED
 A*

Sequence 275
 Contig_0463_pos_18114_19523,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF045058|AF045058_1 Bacillus mojavensis DnaC replicat
 ive helicase (dnaC) gene, partial cds. NID: g3282820.
 40 atggatggaatgtatgagcaaaatcaaatgccgcatagcaatgaagctgaacaatctgtc
 ttaggtgccattattatagatccagaactcattaatactactcaggaagtcttgcttcct
 ggtcggttttatagagggcgcccatcaacatatttttcgagcaatgatgcacctaaatgag
 gataataaagaaattgatgttgtcacattgatggatcaattatcaagtgaaggtagctta
 aacgaagcgggtggccctcaatatctcgccgaactatcgacgagtgtagcgacaacgcga
 45 aatgttcagtactatacggatatcggttttaacatgcgttgaaacggaaacttattcaa
 accgctgatagtagcgaatgatggctataatgatgaattagaattagatacgatttta
 agtgacgccgaacgacgtattttagaactatcttctacaagagaaagtgatggttttaa
 gatattagagatgtcttaggacaggtatatgaaaccgcagaagaactcgaccaaatagt
 ggtcaaacaccaggtattccaactggttatcgtagcttagaccaaatagactgctgggttt
 50 aatcgtaatgatttaattattctagcggcacgtccttcagtaggtaagactgcctttgcc
 ttaaatattgcgcaaaagggttgccacacatgaagatatgtatactgtcggtatcttctca
 cttagagatggcgccgaccaattggcgacacgtatgatttglagttctggtaacgttgat
 tccaatcggttaagaactgggacgatgactgaagaagattggagtcgctttacgattgcg
 gtaggtaagctatcacgaactaaaatcttcatagacgatacgccaggtatccgatatcaat
 55 gatttacgttctaataatgtcgtcgactcaacaagagcagcggcttgatgatgtgtgatt
 gattatctacaattgattcaaggaagcggtacaggtttctcagataaacgtcaacaagaa
 gtttcggagatttcacgtacacttaaggcgattgcacgtgaattagaatgtccagttatt
 gcactgagtcagctatcacgtggcggtgaacagcgacaagacaaacgtcctatgatgagt
 gatattcgtgaatctgggtctatagaacaagatgccgacatcgctcgctttcttgatcgt

gatgattattataatcgtggtgaaggtgatgaagatgatgacgatgctgacgatgctggt
 tttgaaccacagacaaatgatgataacggtgaaattgaaatcatcatcgccaagcagcgt
 aatggtccaacaggtactgtgaaacttcactttatgaaacaatacaataaatttacagat
 attgattatgctcatgcagatatgtcataa

5

Sequence 276

MDGMYEQNQMPHSNEAEQSVLGAIIDPELINTTQEVLLPESFYRGAHQHIFRAMMHLNE
 DNKEIDVVTLMQSLSEAGGPQYLAELSTSVPTTRNVQYYTDIVFKHALKRKLIQ
 TADSIANDGYNDELELDTILSDAERRILELSSTRESDFKDIRDVLGQVYETAELDQNS
 10 GQTPGIPTGYRDLQMTAGFNRNDLIILARPVSFGKTAFAFNIAQKVATHEDMYTVGIFS
 LEMGADQLATRMICSSGNVDSNRLRTGTMTEEDWSRFTIAVGKLSRTKIFIDDTPGIRIN
 DLRSKCRRLKQEHGLDMIVIDYLQLIQSGSRFSNDRQQEVSEISRTLKAIARELECPVI
 ALSQLSRGVEQRQDKRPMMSDIRESGSIEQDADIVAFLYRDDYNNRGEDEDDDDADDAG
 FEPQTNDNGEIEIIIAKQRNGPTGTVKLHFMKQYNKFTDIDYAHADMS*

15

Sequence 277

Contig_0463_pos_19769_0,

is similar to (with p-value 3.0e-77)

>sp:sp|P29726|PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3
 20 .4.4) (IMP--ASPARTATE LIGASE). >gp:gp|M83690|BACADESYN_1 Bac
 illus subtilis adenylosuccinate synthetase (purA) gene, comp
 lete cds. NID: g142442.

gtggttaaacgagaaaaaacttgagggtgctcatatgtcatcaatcgtagtagttgggaca
 caatggggagacgaaggtaaaggtaaaataacagacttttttagcagagcaagcagacgta
 25 attgctagatttttctggtggttaacaatgcgggacatacgattcaatttggtggagaaact
 tacaattacacttagtaccatcaggtatcttttataaaagataaattagcagtaatcggt
 aacggtgtagttgtagatccagtcgcatattataaaagaattagatgggttaaataaacgt
 ggcatttcaactgacaacctacgcacatctcaaatcgcgacacagtcattttaccttatcac
 cttagctcaagacgaatatgaagaacgtcgctcggtggcgataataaaatcggtacaacgaaa
 30 aaaggtattggccagcagcagtagataaaagcacaacgtatcggtattcgcatggcagat
 ttattagaaaaggaaacattcgaacgccgacttaaagaaaatattgaatataaaaatgca
 tactttaaaggcatgtttaacgaaactgtccaacattcgatgaaatctttgacgaatac
 tatgctgcaggtcaacgttttaaagactatgtgacagacacagc

Sequence 278

VVKREKLGGAHMSSIVVGTQWGDEGKGKITDFLAEQADVIARFSGGNNAGHTIQFGGET
 YKLHLVPSGIFYKDKLAVIGNGVVDPVALLKELDGLNERGISTDNLRISNRAQVILPYH
 LAQDEYEERRRGDNKIGTTKKGIGPAYVDKAQRIGIRMADLLEKETFERRLKENIEYKNA
 YFKGMFNETCPTFDEIFDEYYAAGQRLKDYVTDTA

40

Sequence 279

Contig_0463_pos_18985_18680,

is similar to (with p-value 2.0e-31)

>gp:gp|AF045058|AF045058_1 Bacillus mojavensis DnaC replicat
 45 ive helicase (dnaC) gene, partial cds. NID: g3282820.

atgaagatttttagttcgtagatgcttacctaccgcaatcgtaaagcgactccaatcttct
 tcagtcacgtcccagttcttaaacgattggaatcaacgttaccagaactacaaatcata
 cgtgtcgccaattggctcgccgcccattctcaagtgagaagataccgacagtatacatatct
 tcatgtgtggcaaccttttgcgcaatatttaaggcaaggcagtccttacctactgaagga
 50 cgtgccgctagaataattaaatcattacgattaaaaccagcagtcatttgggtctaagtca
 cgataa

Sequence 280

MKILVRDSLPTAIVKRLQSSSVIVPVLKRLESTLPELQIIRVANWSAPISSEKIPTVYIS
 55 SCVATFCAIFKAKAVLPTEGRAARIKSLRLKPAVIWSKSR*

Sequence 281

Contig_0463_pos_11710_10706,

is similar to (with p-value 2.0e-89)

>sp:sp|P09978|PHLC_STAAU PHOSPHOLIPASE C PRECURSOR (EC 3.1.4.3) (BETA-HEMOLYSIN) (BETA-TOXIN) (SPHINGOMYELINASE). >pir:p
ir|S15766|S15766 beta-hemolysin - Staphylococcus aureus >gp:
gp|X13404|SAHLB_2 Staphylococcus aureus hlb gene for beta-he
5 molysin. NID: g46586.
atgaaacgaggtgtaacaatattgaattggcaacgtaaatgtataactactcttgttg
gttttaagtagtttatttttagtattttcgactatcacatatgcgagtgaacgtgatttt
aaagacagtccttaaaatcactacacataacgtgtattttcttacctactgctatctaccct
aattggggacaatctcagcgcgctgatttaatttcaaaagcagattacattcaaaatcaa
10 gatgtcgtgattctaaatgaattatttgataaaaaagcttcaaaaagattgttaacacgt
ctacattcacagtacccttatcaaacacctatcggttggttaaaggtagagaaggttggcaa
aatacttctggtacttatagaaaaattaaaaaagtaagtggtggcgttggtattgtgagt
aaatggcctatcgtaacaagaacaacatatattataaaaaaggttggtgggctgatatg
gcaggtaagtaaaaggcttgctacattaaaattaataagaatggcaaataccaccatatt
15 atcggaacacatctacaagctgaagatccaacatgctttaaaggaaaagataaagacatt
agacagagtc aaatgagtgaattaaacagtttatcaaagacaagaatatccctaaaaat
gaaccgcgtctatatcggtggtgacttaaatgtcattaaagattcagatgaatatcaacaa
atggcaataaacttaaatgtttcattacctactcaattcgatggtaatgcataatggttg
gatactagcagtaaatgatttgcgaaatataattatcctaaattagaacctcaacactta
20 gattatattttattagatcatgaccatgcacaaccaagctcatggcataatgatacacat
agagtgaagtcaccagaatggtccgtgaaatccttggggaaaaacatacaaatacaatgat
tactcagatcattaccactctcaggctatgcataaatgaatag

Sequence 282
25 MKRGVTILNWQRKCILTTLLVLSSFLVFSTITYASERDFKDSLKITTHNVYFLPTAIYP
NWGQSQRADLISKADYIQNDVVILNELFDKASKRLLTRLHSQYPYQTPIVGKGTEGWQ
NTSGTYRKIKKVSGGVIVSKWPIVQQEQHIYKKGCGADMAGNKGFAIYIKNKNGKYHHI
IGTHLQAEDPTCFKGKDKDIRQSQMSEIKQFIKDKNIPKNEPVYIGGDLNVIKDSDEYQQ
MANNLNVSLPTQFDGNAYSWDTSSNSIAKYNYPKLEPQHLDYILLDHDHAQPSSWHNDTH
30 RVKSPKSVKSWGKTYKYNDYSDHYPLSGYASNE*

Sequence 283
Contig_0463_pos_4418_3897,
is similar to (with p-value 7.0e-52)
35 >sp:sp|P12012|GNTP_BACSU GLUCONATE PERMEASE. >pir:pir|A26190
|A26190 gluconate permease - Bacillus subtilis >gp:gp|AB0055
54|AB005554_2 Bacillus subtilis genomic DNA, 36 kb region be
tween gnt and iol operons. NID: g2280496. >gp:gp|J02584|BACG
NT_3 B.subtilis (gluconate operon) gntR, gntK and gntP genes
40 encoding gnt repressor, gluconate kinase and permease, and
gntZ gene. NID: g143013. >gp:gp|Z99124|BSUB0021_112 Bacillus
subtilis complete genome (section 21 of 21): from 3999281 t
o 4214814. NID: g2636442.
atgcttatcgcagtgatatttgcaatctttacaatgggaatgaagcaacaacggaaaatg
45 gaagacattatgaaatcagttacgcatgctatttatccaatcggcatgatgttactcatc
atcggtggtggtggtacatttaacaagtgctcatcgatggtggcgttaggtgatacaatc
gctaagatgtttgaaggacaagcatgtcgccattttattagcatggattgtagctgca
gtcttaaggattttcattagatcagctacagttgctgccgtatcaacaacaggcattgtg
ttaccacttttagaacattcagatgttaaatgtagctttggtcggttcttgcaataggtgca
50 ggtagcgttaattctctctcacgtcaatgatgctggattctggatgttttaagaatatttc
gggctgacagttaaagaaacatttttaacatgggtcggtatttagagacaattatttcagta
tctggtattttattttattttatttatcagtttatttgtag

Sequence 284
55 MLIIVIFAIFTGMKQQRKMEDIMKSVTHAIYPIGMMLLIIGGGGTFKQVLIDGGVGDIT
AKMFEGTSMSPILLAWIVAVALRISLGSATVAAVSTTGIVLPLEHSDVNVALVLAIGA
GSVILSHVNDAGFWMFKEYFGLTVKETFLTWSLLETIIISVSGILFILFISLHV*

Sequence 285

Contig_0463_pos_3576_2782,
is similar to (with p-value 0.0e+00)
>sp:sp|Q05852|GTAB_BACSU UTP--GLUCOSE-1-PHOSPHATE URIDYLYLTR
ANSFERASE (EC 2.7.7.9) (UDP-GLUCOSE PYROPHOSPHORYLASE) (UDPG
5 P) (ALPHA-D-GLUCOSYL-1-PHOSPHATE URIDYLYLTRANSFERASE) (URIDI
NE DIPHOSPHOGLUCOSE PYROPHOSPHORYLASE). >pir:pir|A40650|A406
50 UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)
- Bacillus subtilis >gp:gp|L12272|BACGTABX_1 Bacillus subti
lis UDP-glucose pyrophosphorylase (gtaB) gene, complete cds.
10 NID: g289286. >gp:gp|Z22516|BSLYTGTA_3 B.subtilis lytR, orf
X, and gtaB genes. NID: g405620. >gp:gp|Z99122|BSUB0019_64 B
acillus subtilis complete genome (section 19 of 21): from 35
97091 to 3809700. NID: g2636029.
atgccaaaagaaatgttaccatattagataaaaccaacaattcaatatattgtagaagaa
15 gcttttaaatgcaggaatagaagatattattatagtactggcaagcataaacgtgcaatt
gaggatcactttgacaatcaaaaagaactagagatagtacttgaaagtaaaggaaaagca
gatttacttgaaaaagtacaatattcaacagatttagctaataattttttacgtgcgacaa
aaagaacaaaaagggttaggacatgcaattcactactgcaaaacagtttataggtaacgaa
ccatttgcagtggttattaggagatgacattgtagagctgatacaccagctattaaacaa
20 ttaatggatggtttatgaagaaacaggccattcagtaataagggttcaagaagtgccagaa
tctgatacacatcgctatggtgtgattgatccttctgctaaagagggaagtcgatatgaa
gtacgtcaattttagataaaaagccgaaacaaggtactgccccgtctaatttagcaatcatg
ggtcggttatgtattaacaccagaaatttttgattatcttgaaacacaacaagaaggtgct
ggaaatgaaattcaattaactgatgcgattgaacgaatgaatagcaacaacaagtgat
25 gcatatgattttgagggtaatcggttatgatgttgagaaaaattagggtttgttaaaaca
acgattgaatatgctttaaaagatccagaaatgagtcaagacttaaaagcattcattaaa
caactagatatatttaa

Sequence 286
30 MPKEMPLPILDKPTIQYIVVEEAFNAGIEDIIIVTGKHKRAIEDHFDNQKELEIVLESKGKA
DLLEKVQYSTDLANIFYVRQKEQKGLGHAIHTAKQFIGNEPFAVLLGDDIVESDTPAIKQ
LMDVYEETGHSVIGVQEVPESDTHRYGVIDPSAKEGSRYEVRQFVEKPKQGTAPSNLAIM
GRYVLTPEIFDYLETQQEGAGNEIQLTDAIERMNSKQQVYAYDFEGNRYDVGEKLGFFVK
TIEYALKDPEMSQDLKAFIKQLDI*

Sequence 287
Contig_0464_pos_1580_2050,
is similar to (with p-value 4.0e-35)
>sp:sp|Q02134|HIS7_LACLA IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRA
40 TASE (EC 4.2.1.19) (IGPD). >pir:pir|G45734|G45734 HisB - Lac
toccoccus lactis subsp. lactis >gp:gp|U92974|LLU92974_6 Lacto
coccus lactis unknown gene, partial cds, and HisC (hisC), un
known, HisG (hisG), unknown, HisB (hisB), unknown, HisH (his
h), HisA (hisA), HisF (hisF), HisIE (hisIE), unknown, unknow
45 n, LeuA (leuA), LeuB (leuB), LeuC (leuC), LeuD (leuD), unknow
n, IlvD (ilvD), IlvB (ilvB), IlvN, IlvC (ilvC), IlvA (ilvA)
, AldB (aldB) and aldR (aldR) genes, complete cds. NID: g256
5137.
atgttaacgctatttacttttcatagtggattaactttatctattgaggccactggagat
50 acgtatggtgatgatcatcatataactgaagatataggtatagttattggacaattactt
cttgaattaataaagactcaacaaagttttacaagatatggttgctcatatgtacccatg
gatgaggcgcttgcgcgaacagtagtgacattagtggtcgtccatatttctcattta
atgaagattgagcgctcaaaaggttaggaacttttgacactgaactagttgaagaattttt
agacattgataaataatgcgcgattaacccgttcacattgacttattaagaggtggaat
55 acacatcatgagattgaggcaatatttaaatcttttgcaagagcattaaagatttctctt
gcacaaaatgaagatggacgtattccatcgtctaaaggagtaattgaatga

Sequence 288
MLTLFTFHSGSLTSLIEATGDTYVDDHHITEDIGIVIGQLLLELIKQTSFTRYGCSYVPM

DEALARTVVDISGRPYFSFNSKLSAQKVGTFDELVEEFFRALIINARLTVHIDLLRGGN
THHEIEAIFKSFARALKISLAQNEGGRIPSSKGVIE*

Sequence 289

- 5 Contig_0464_pos_2149_2625,
is similar to (with p-value 1.0e-29)
>sp:sp|Q02132|HIS5_LACLA AMIDOTRANSFERASE HISH (EC 2.4.2.-).
>pir:pir|I45734|I45734 HisH - Lactococcus lactis subsp. lac
tis >gp:gp|U92974|LLU92974_8 Lactococcus lactis unknown gene
10 , partial cds, and HisC (hisC), unknown, HisG (hisG), unknow
n, HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (his
F), HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB
) , LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilv
B), IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (al
15 dR) genes, complete cds. NID: g2565137.
gtgcaaaaagctgaagctatcgctacttccaggtgttgacattttcaggatgcgatgcat
tctatagaagaaaaagcattaaagatatgcttaaaaaatatacatgataaaccgataatt
ggaatatgtttaggtatgcaattactttttcaacatagcgcagaaggtgacggttagtgga
ttggaacttgctcccggaatatagtgccaatccaatcatctcatcctattcctcatttg
20 ggttggaatgaattaaagagtacacatcccttactgcaaagtgatgtgtattttgttcat
tcatatcaagcagaaatgtcagaatatgtagtagcttatgctgactatggtacaaagatt
ccgggagtcattcaataaccgaaattatataggtatccagtttcatcctgaaaaagtgga
acgtatggattagagattctaaatcaagcgcttaaggagggtttattaatgattga

25 Sequence 290

VQKAEIVLPGVGHFQDAMHSIEEKSIKMLKNIHDKPIIGICLGMQLLFQHSAGEDVSG
LELVPGNIVPIQSSHPIPHLGWNEKSTHPLLQSDVYFVHSYQAEMSEYVAYADYGTKI
PGVIQYRNYIGIQFHFPEKSGTYGLEILNQALKGGFIND*

30 Sequence 291

- Contig_0464_pos_3334_4077,
is similar to (with p-value 5.0e-69)
>sp:sp|O34727|HIS6_BACSU HISF PROTEIN (CYCLASE). >gp:gp|Z991
21|BSUB0018_173 Bacillus subtilis complete genome (section 1
8 of 21): from 3399551 to 3609060. NID: g2635827. >gp:gp|AF0
35 17113|AF017113_41 Bacillus subtilis 300-304 degree genomic s
equence. NID: g2618830.
gtgattccatgttttagatgttaaagatggacgcgctgtaaaggggtatccagttccagtc
ttaagagatatcggtaatccagttgatttggctctttattataatgaagccggtgcagat
40 gaactagtctttcttgatatttcgaagacggaagcaggacatgatcttatgatagaagt
atagaagcaacggcaaaaacaattatttatccctttaacagtaggaggaggattcaaaat
ttagatgatattacgcaactattaaatcacggagcagataaaatatactcaattcaagc
gctttaaaacatccagaattaattcgacaagcaagcgagaaaatttggtcgtcaatgtatt
tgtattgctattgtagcttttatgataaagacagaaaggattatttctgtactacgcac
45 ggtggtaaaaaattaactgatgtcagggtatatgattgggtacaagaagtagagctttta
ggtgctggggaattgcttataactagcatgcatcatgatggaatgaaacaagggtttgat
attgaacatttagcaaaaattaacaattagtttaattccgattattgcctctgggggt
ggaggaaatgcacaacattttgttgaaactatttcaacaacagatgtttcggcagggtta
gcggcaagtattttacatgatcaagaaactacagtggcagaaattaaagataaaatgcgt
50 gaaggagggtatccttgtgagatga

Sequence 292

- VIPCLDVKDRVVKGIQFQSLRDIGNPVDLALYYNEAGADELVFLDISKTEAGHDLMIIEV
IEATAKQLFIPLTVGGGIQNLDITQLLNHGADKISLNSSALKHPELIRQASEKFGRQCI
55 CIAIDSFYDKDRDYFCTTHGGKKLTDVRVYDWVQEVLLGAGELLITSMHHDGMKQGF
IEHLAKIKQLVNIPIIASGGGGNAQHFVELFQQTQDVSAGLAASILHDQETTVAEIKDKMR
EGGILVR*

Sequence 293

- Contig_0465_pos_9467_9787,
putative peptide of unknown function
atgtaccggcgatgggtatctttttcaactacagctacttgcttaccacttttgctttaatg
ttaacgcagcatgccaaagctgcatgaccacttctaaaaatactacatcatattgtttca
5 ttaacactcatcctttcttatttttctatgagatgttttaatgtttgctctagttcttca
aacacatatctcgtttcatcactcagtcgtatcaaattcttggtaagcaacttgaa
attgcttcaaaaaacagcttcttgttgttggccattgtcagttaacgtaattatcaac
tgtcgtttatcagattgttga
- 10 Sequence 294
MYRRWYLFQLQLLAYHFALMLTQHAKLHDHFLKILHHIVSLTLILSYFSMRCFNVCSSSS
NTYFVSSYSVVSNSCGKQLEIASKTASCCCPVSVNVIINCRSLDC*
- Sequence 295
15 Contig_0465_pos_11617_11937,
putative peptide of unknown function
atgagtgataatacaccaccaataaagaacatcctcattacatcaaagataactaatgttt
ttaaatgcattagattcgaagaagaagattaatcctgaaataggaactaatagtgcgcca
ataaatatcatacctgggtatggcattggcattaccaaataatattgttaatatccataac
20 tctaggaaagtgtatcctaaagctaaaaatacacgtgagaacaccaaggctctcccccac
tctcggaaacttcattaatatgtggcgctgtacgttttgttccagcaataatacatca
tccgcttcatctttggtgtga
- Sequence 296
25 MSDNTPPIKNILITSKILMFLNALDSKKKINPEIGTNSAPINIIIPGMALALPNIFVNIHN
ARKVIPKAKNTRENTQGLPHSSETSLICGVVRFVPAINTSSASSLV*
- Sequence 297
Contig_0465_pos_15548_16303,
30 putative peptide of unknown function
atgtttaaagttagttatattgtgatgatgaaaggattataagagaaggcttaaagcaaatg
gttccatgggaggactatcatttcaccactgtttatactgccaaagacggcggtggaagca
ttgtctttaattcgccaacatcaacctgaactcgctcattactgatatacgaatgcctcga
aaaaatggtgttgacctactagatgacatcaaagaccttgattgccagattatcatttta
35 tcgagttatgacgacttcgaatatatgaaagccggtatacaacatcatgttcttgattat
ttactaaagccagtagaccacactcagtttagagcatattctagacatattagttcaaagg
ttattagaacgcccacattctaccaatgatgacgcgcatatcactgcctttcaacca
ttattaaaaattgattacgatgactattatgtcaatcaaattttgtctcaaataagcaa
cattatcacaaagaaagtgactgttcttgacttaattaatcctattgatgtaagtgaagca
40 tacgccatgaggacgttttaagaacatgtaggcattacgatagttgattatcctaaatcgt
tatcgtattttaaaatcattacatcttttagaccagcactacaagcattatgaaattgct
gaaaaagtaggtttttctgagtataaaatgttttgcctatcattttaaaaaatatttacat
atgtcaccaagtgattataataagcaatcaaaatag
- 45 Sequence 298
MFKVVICDDERIIREGLKQMPWEDYHFTTVYTAKDGEALSLIRHQPELVITDIRMPR
KNGVDLLDDIKDLDCQIIILSSYDDFEYMKAGIQHHVLDYLLKPVDHTQLEHILDILVQR
LLERPHSTNDAAAYHTAFQPLLKIDYDDYYVNQILSQIKQHYHKKVTVLDLINPIDVSES
YAMRTFKEHVGITIVDYLNRIRILKSLHLLDQHYKHIEIAEKVGFSEYKMFYHFKKYLH
50 MSPSDYNKQSK*
- Sequence 299
Contig_0465_pos_14779_13595,
putative peptide of unknown function
55 atgcaaacagtcggaattataccttcgccaggtatagcacatcaacatgcaaaaaaata
attccaaatgttaaacagttattgtcaaagcgtagtaaacatagtcaatggaatttcgac
atcaaagtcgatctcatgataggatctgcagaggatgtacatgaaagtgttgaaaaagca
gcacaaattaaagaggaacatcagtgaggattacgttgtttgtctgacagatttgcctagt
atttcagataataaagtgggtgtcagcgactttaatagtgcacaaacatgttgcaatgcta

tcattaccgtcactaggtttttattgatttgaagcgcaagctagttaaaacgatgacttca
 ttgattgaacaattatattataatcaaccgaaagacaaaaatgcgccacatccttttgta
 cgcgtgaaggctgtagaacctgacgaagacgccacatcaaaacaacgatataatatt
 tttttatcataagtggattcagtttaattgggtggactgacacgagcaaatcagccttg
 5 aaaaacatctttaattttaagaaaaatcatttcagttgcctttgcaacaggaacttatgtc
 tcaatattttcaatgccatgggaattaagcgtgatttattcaccgcttcgacttatcata
 ttgatgggtgattgctatacttgggatggctggatggctatttctatgcgcatcaattgatt
 gaaaagaaaactgctaaatctcagcgtgtatatcgatatatttataaattcaaccacactt
 gttacactaagtttgattacactcataaattatgtcattttataatttattgttaatcatc
 10 agtattacactcctttgtccctgtggaattatttaatagttggacgagtgcccaatcacia
 tttacgttctcaaatatatgagattgatttgggttgtatcatcatttaggacttttagct
 ggagctatgggatcaactggtgaaaatgaagagaaaatacgtcgtattacttattcttat
 agacaatatcatcgttataaagaagctgagcaagaacaaaaagaacaagaaacttctcgt
 gatgtatcacacaaaatgtcgaacaacaaacttcaagtaagatgaaaataatgaacaa
 15 tatgaaggtaaaaaacaaggacatagagaggaggatgacgcatga

Sequence 300

MQTVGIIIPSPGIAHQHAKKIIIPNVKQLLSKRTKHSQWNFDIKVDLMIGSAEDVHESVEKA
 AQIKEEHQWDYVVLCLDLPISIDNKVVVSDFNDSKXVAMLSLPSLGFIDLKRKLVTMTS
 20 LIEQLYYPKPKDNAPHFVRVKAPEDEATSKQRYINILFIISWIQLIGGLTRANQPW
 KNIFNFKKIIISVAFATGTYSVIFSPWELSVIYSPRLRIILMVIAILGMAGWLFYAHQLI
 EKKTAKSQRVRYIYNSTTLVTLISLITLINYVILYLLLIISITLFVPVELFNSWTSQSQ
 FTFSNYMRLIWFVSSLGLLAGAMGSTVENEEKIRRTYSYRQYHRYKEAEQEKEQETSR
 DVSQONVEQQTSSKDENEQYEGKKQGHREDDA*

Sequence 301

Contig_0465_pos_13580_12489,

putative peptide of unknown function

gtgggtctagtcgctcaggtgttactgaacgccttgagaaaaatctcatacaagaa
 30 atgcctaaaaatgttatctacgcattatgatcatcagcaagaatggatttttgatttagtt
 actgatccgcttactgggtttgctgaatctgtagatgaaatttttgagaaagtagccgat
 tatcacgataagagacaatgggattatgtgatagcaattacagatttaccgatgtttgct
 gacaaacaagtgatggcattagatattaatattgaaaatgggtgcagctatattctcatat
 ccggcattttggctggcgtccagtaaaaaaacgtttcaagcatgcgatttataatattatt
 35 caagaattaaatgaagctgaacaagaaagtcgtaattatgataataataatcaaatagaa
 aattcagtaaaaaaacaatttccgctctctaaaaatagacaaagaaacaatatatatgaaa
 gaaacagactcttatcacttaagatatttatcaagttcacgttctagaggcatgtttcgc
 cttgttagtggaatgacattttgcgaataatccattaaatatgatggcaagtttaagtaat
 atagtagctattgcatttactacaggtgcattttggacttgtatttacaacgatgtggcaa
 40 atggcttataacttttcaatgtggcgtttattttggaatttcaattattgcgattattgga
 atgctaataatggataatgatgtcacatgatttatgggaaccagttaataaaagcaaccat
 aagcgtattacttgggtatacaatcttacaacaataatgacattgatttttgccattata
 atttattatattattctttatttactattcttaattgctgaaatcgattattgcatca
 ggatttttaggtcagcaagttggattaaaaggtcctgcaggcattgatttatatttaagt
 45 attccatggtttgcagcttcaatttcgacagttgcaggtgcaatagggtgctggtttactt
 aatgatgaactcattaaagaaagcacatatggatatcgtcagcgtgtaagatacgaagaa
 caacgtcgataa

Sequence 302

VGLVVPAGVTERLAENLIQEMPKMLSTHYDHQEWIFDLVTDPLTGFAESVDEIFEKVAD
 50 YHDKRQWDYVIAITDLPMFADKQVMALDINMENGAAIFSYPAFGWRPVKKRFKHAIYNI
 QELNEAEQESRNYDNNNQIENSVKKQFPLSKIDKETIYMKETDSYHLRYLSSSRSRGMFR
 LVSGMTFANNPLNMASLSNIVAIPTTGAFLVFTTMWQAMAYNFSMWRLFGISIIAIG
 MLIWIMMSHDLWEPVKNHNRITWLYNLTTIMTLIFAIIIYIIILYLLFLIAEIVLLPS
 55 GFLGQQVGLKGPAGIDLYLSIPWFAASISTVAGAIGAGLLNDELIKESTYGYRQVRVYEE
 QRR*

Sequence 303

Contig_0465_pos_11648_11019,

putative peptide of unknown function

atgttctttattggtggtgtattatcactcattagtagacaatgattttatatcaatttggt
acatttagtactgaatcgcaatattatggaattatgactataacagatgcgttcataagta
ggatttggtgaagagcttgaaaagcaactggtgtattttatttattaattatttataaa
5 acaataaaaatactcaatggattacttatcggtgctgctgttggtgcagggttcgcggtg
ttgaatcagctggttatatctttagggttggttaattttatttgatggagtaataat
attactgaaactactatacaaaagaggttgacagctttaggtagtcacctcgtttgggca
gctatttggtgctgctgcggcagtaatagtgaaagaaacaaagcatttcgaatgggcaat
atcatcgataaacgttttatattttctttttgtggcagtgacattacacggaatatgg
10 gatacggaaataacacttttaagtagtggttattttaaataatatcttattaattatgatt
gcatggctattttatatcttatacttatgaaagcagggttaactcagggtgaatcagttgcgt
gatgaatacaatcgttttagaggaaaggtga

Sequence 304

15 MFFIGGVLSLISTMILYQFVTFSTESQYYGIMTITDAFIVGFVEELGKATVVILFINYLK
TNKILNGLLIGAAVAGAGFAVFESAGYIFRFGFNLFDGVNNITEITIQRGWTALGSHLVWA
AIVGAAAVIVKETKHFEWANIIDKRFIFFFVAVTLHGIWDTEITLLSSGYLKYILLIMI
AWLFIFILMKAGLTQVNQLRDEYNRLEER*

20 Sequence 305

Contig_0465_pos_11013_10264,

putative peptide of unknown function

atgaaattctgccctcattgtggaatccgataaaaaaggaaacagtcattttgtaataaa
tgtggaataacattttaaagacatcgacacaaagaaaaagtgaaatcaaattgaacatatg
25 cgtgaacagcaatcgtatatttcttgtgaggaaagacaacatcatgattcaacattttat
aaagaacaaaaacatactgggtggctaattgtattatcaattatatttgccttggtgata
gcagcgtatttgatggtgctactatacttacaatcattatatttagtgatgagcaaaagt
catcaaacacacagctctcagcaatcaaagtggtcaaaataaggatcaatccact
ggtccaagcattgatgttttttagtgatgactttgatcaagggttatatgaagtcagcttca
30 acaagtggtatagaggtgtttataatggaatgacacgtgaagaagttgaagataaattt
ggaacatccaatggttctgtagaaagttgaagtgagttacgaaaaatatggtgattta
gctgtagcctacgatgataatgaagttgttagcgtaggtgtagcacctaatacatatttca
gaagatcaatttttaagtagtatataatgaaccggatgatagaaattcaagccaactcatt
tatgatagtaacaaagataatgacttctctgtgttagctaattgttaaaatggatatgtt
35 actgtcattgaaaaatgtaaatcaaatttaa

Sequence 306

40 MKFCPHCGNPIKKEQSFNCNKGKHLKTSTQRKSENQIEHMRQQSYISCEERQHHSTFY
KEQKHTGWLIVLSIIFVLLIAALLYGAYTYNHYISDEQSHQTTQSQSNESGQNKDQST
GPSIDVFSDDFDQGYMKASSTSGYRGVYNGMTREEVEDKEGTSNGSVESLKWSEYKYGDL
AVAYDDNEVSVGVAPNHISEDQFLSMYNEPDDRNSSQLIYDSNKDNDFSVLANKNGYV
TVIENVNQI*

45

Sequence 307

Contig_0465_pos_7177_6710,

is similar to (with p-value 8.0e-58)

>gp:gp|AJ000974|BSPYREYLO_2 Bacillus subtilis pyrE to yloA g
50 ene region. NID: g2462954. >gp:gp|Z99112|BSUB0009_28 Bacillu
s subtilis complete genome (section 9 of 21): from 1598421 t
o 1807200. NID: g2633902.

gtgaaagataaatatccgcaattacgcattaaaatgaaaaaacggaaacttacgttagag
gaacaaggtgagaaatataatcctgctttatggaagaatgatcctaaccaatgttgctac
55 atacgcaagattaaaccactagaagacgtattatctggtgctgtagcttgatcagggt
cttagacgagcacaatcaccaacacgagcacatacaaatctcatttaacaaagatgaaaga
tttaagtcattaaagtggtgctcccttaattctattggacagaagaagaagtaggtcttat
atacgtgataaggatttaccatataatgaattacatgatcaaaattatccaagtaggtggt
tgtattccatgtacatcaccgctatttgattctaattgattcacgtgctggtcgttggtcc

aattctagtaagactgaatgtggattacatgtagctgataaaccataa

Sequence 308

VKDKYPQLRIKMKPELTLEEQGEKYNPALWKNDPNQCCYIRKIKPLEDVLSGAVAWISG
5 LRRASQSPTRAHTNFINKDERFKSIKVCPLIYWTEEEVWSYIRDKDLPYNELHDQNYPSIG
CIPCTSPVFDSDNSRAGRWSNSSKTECGLHVADKP*

Sequence 309

Contig_0465_pos_2817_2065,
10 is similar to (with p-value 1.0e-71)
>sp:sp|P29928|SUMT_BACME UROPORPHYRIN-III C-METHYLTRANSFERAS
E (EC 2.1.1.107) (UROGEN III METHYLASE) (SUMT) (UROPORPHYRIN
OGEN III METHYLASE) (UROM). >pir:pir|A42479|A42479 S-adenosy
l-L-methionine uroporphyrinogen III methyltransferase - Baci
15 llus megaterium >gp:gp|M62881|BACCOBA_1 Bacillus megaterium
S-adenosyl-L-methionine:uroporphyrinogen III methyltransferas
e (COBA) gene, complete cds. NID: gl42694.
gtggttcataatgggaaagtatatatttagttggagctggacctgggtgatccagaattaata
acgttaaaagggttttaaagccattaaagaagccgatgtcatcctttatgaccgacttgta
20 aataaagaaataacttaattatgcttctccttctactaagtcttcttattgaggtaaggat
cctcacaggcactccttacgcaggaagaacaaataaaatgatggtaaccttagccaaa
aaagggcacatagttacacggtttaaggggtggcgatccatttggttttgacgtggcgga
gaagaagcagaggaattagcatgtcataatatccactttgaaattatacctggaattcca
gtaacacatcgtagttatagttcttctgtagcatttgtaactgcagtgaataaacctggt
25 atggataaaggcaataactggcaacatttgcccaatggctcctgaaactttatgtatttat
atgggggttaagagactcagtgaaatttgtaggttgtaatacaatatggctcggttcgtca
gaaacaccagtagctctcgtagcatatgggaacgtcaaaacagcaaatgacagtgactggg
acactcgatacaattcaagaacgagcacatcatattcagaatccagcaatgattattgta
ggcgaagtgggttaagatgagagaaaaaatttaattggtttgtagaacaggcaactgttcaa
30 aatgaaacggttaacggaaatgtcatcaacttag

Sequence 310

VVHMGKVYLVGAGPGDPELITLKLKAIKEADVILYDRLVNKEILNYASPSTKFFYCGKD
PHRHSLPQEETNKMVMVTLAKKGHIVTRLKGGDFVFGRGGEAEELACHNIHFEIIPGIP
35 VTHRDYSSSVAFVTAVNKPGMDKGKYWQHLANGPETLCIYMGVKRLSEICELLIQYGRSS
ETPVALVHMGTSKQQTMTVTGTLDTIQERAHHIQNPAMIIVGEVVKMREKINWFVEQATVQ
NETLTEMSS*

Sequence 311

Contig_0465_pos_1984_1379,
40 is similar to (with p-value 4.0e-19)
>gp:gp|AJ000974|BSPYREYLO_8 Bacillus subtilis pyrE to yloA g
ene region. NID: g2462954. >gp:gp|Z99112|BSUB0009_34 Bacillu
s subtilis complete genome (section 9 of 21): from 1598421 t
45 o 1807200. NID: g2633902.
atgcccttaatgattgatttaagtaacaagaaagtcgctcattgtaggtggaggtaaagt
gcaacacgctcgctgctaaaactttattagcttatacaaaacatattcatgttgtaagtcca
acaattaccgatacattacaaaaatatctagaaacgaagcaaatcacttatgaaaagaaa
cacttcgaaccacaagatggttgagaatgctgatgtggtcatcgcggtactaatcaatct
50 gatgttaacaacgatgtggggcgagctttgtctaaagaacgtattatttaacatgcagga
caagcagacctaggttaagttaacggtccctaatttcttaaaaagagataaattaacaata
agtgtatcaactgatggtgcaagtcctaaattaggtcaacgaattattaaagatttaaaa
gatacatacaatgaagactattcaatgtatattcagtttttatatgaaagtagacaatat
attaaatcacttaaaattgagccatctgataaacaagcgttactcgagcaaattttgta
55 gacaaatatttagatgagaagaagcaacaagatttcacccgatggctaaaaatcacaagtc
aatga

Sequence 312

MPLMIDLSNKKVIVGGGKVATRRAKTLLAYTKHIHVVSPTITDTLQKYLETKQITYEKK

HFEPQDVENADVIAATNQSDVNNVDVGAALSKNVLFNHAGQADLGNVTFPNFLKRDKLT
SVSTDGASPKLGQRIIKDLKDTYNEDYSYMIQFLYESRQYIKSLKIEPSDKQALLEQILS
DKYLDEKKQQDFIRWLKSQVK*

5 Sequence 313

Contig_0465_pos_1292_459,

putative peptide of unknown function

atgggatttggcgcttcacgtcatcaatattattaacttacggtatagcaccggcagta
gtgtcagcaaccgttcatttttctgaaattgcaacaacagctgcatctgggacatcacat
10 tggagatttgataatgttcataaaccaacaatgttgaagttagctatacctgggtcaata
agcgccctttatcggtgcaggtgttttgacatttattcatggtgattatattaaaccattc
attgctttattcttgttaagtattgggattttatattttgtatcaatttctatttaaaccgt
gcacatgaacatcatcatcatgtgggaaatttgagtagttttaaagtaattccacaaggt
tttgtggcaggatttttagacgcaatcggtggtggtggtggggaccggttaatacgcgcg
15 ctctgctttcaagtaaaaaaattcaaccacgatatgcgattggaacagctctcagcaagt
gaattttttgttacgtcatctgccgctttaagtttcattatcttttttaggagtcactcaa
attaattggtttgctgtaattgctttaagtcctcggtggaatggtagcagcacctatttca
gcgtatttagttaaagtgttaccatttaacattcttgcaatttgtgctcggtggtttaatt
atttttacaaatagtaatgcattattaagctattttgtaaaagataaacactatttcaa
20 acagttcgattcattattattcttgcaattattattttgcttggttttcaagtcggtcga
aacaagaaattgtctttttcttataagaaaagccgagtaaaacaatatataattaa

Sequence 314

MGFGASSSSILLTYGIAPAVVSATVHFSEIATTAASGTSWFRFDNVHKPTMLKLAIPGSI
25 SAFIGAGVLTFIHGDYIKPFIALFLLSMGFYILYQFLFKRAHEHHHHVGNLSSFKVIPQG
FVAGFLDAIGGGWGPVNTPLLLSSKKIQPRYAIGTVSASEFFVTSSAALSFIIFLGVTQ
INWFAVIALSLGGMVAAPISAYLVKVLPI NILAICVGGIIIFTNSNALLSYFVKDNTISN
TVRFIIILAI IILLVFQVVRNKKLSFSYKKS RVNKYN*

30 Sequence 315

Contig_0465_pos_0_435,

is similar to (with p-value 7.0e-26)

>gp:gp|AJ000974|BSPYREYLO_4 Bacillus subtilis pyrE to yloA g
ene region. NID: g2462954. >gp:gp|Z99112|BSUB0009_30 Bacillu
35 s subtilis complete genome (section 9 of 21): from 1598421 t
o 1807200. NID: g2633902.

atgtctaacaatgaaacaataaccaattatacaattaaacctcatggaggagaactcatc
aatcggtgtgttgaaggaaacgaacgtgaacgtttgattgaggaagcattaaatttttaa
ccgattacttttaaatccttggggaatatcgatctagagctcataggtattggcggattt
40 agtccccttacaggatttatgaacaaggaagactacactaaggttatagaggaaacacat
ttaagcaatggcttagtttgaggtattcctatcactttacctgtaacagaatccgaagca
gataaacttgaaataggtgatgatattgctttatatggtgaagatggtcagttatatgga
acgcttaaattagaagaaaagtaacatatgataaagaaaaagaagcgcgtttggtgtac
ggaactactgaagaa

45

Sequence 316

MSNNETITNYTIKPHGGELINRVVEGNERERLIEEALNFKPITLNPWGISDLELIGIGGF
SPLTGFMNKEDYTKVIEETHLSNGLVWSIPITLPVTESEADKLEIGDDIALYGEDGQLYG
TLKLEEKYTYDKEKEARLVYGTTEE

50

Sequence 317

Contig_0466_pos_3615_2260,

is similar to (with p-value 0.0e+00)

>gp:gp|Y09570|SAFEMD_1 S.aureus femD gene. NID: g1684748. >g
55 p:gp|Y15477|SAARGFEMD_4 Staphylococcus aureus argI, glmM gen
es and ORF1 and ORF2. NID: g3892891.

atgggaaaatttttgggtactgatggtgttcggtgtcgctaaccaagaactcacacct
gaattggcttttaactaggttagatagcaggagatattgttctcgacataataagggtgaa
aagcatcctcgagtttttagtaggaagagatacaagatttcaggagaaatgctagaatct

gcattaattgctgggttaatttcaattggcgcagaagtgatgcgcttaggtgttatttca
 acaccgggtgtggcttatttgactaaagaaatggaagcagcattaggtgttatgatttct
 gcgtcacataatccggttgctgataatggaattaaatttttgggttcagatggctttaa
 5 ttgtcagatgatcaagaaaatgaaattgagcaattattagatcaaaccaatcctgattta
 ccacgaccagttaggagaggatattgtacattattcagattattttgaaggtgcacaaaag
 tatctaagttatcttaaatcaactgttgatgttaattttgagggctttaaattgtatta
 gatgggtgcaaacgggtcaacttcttcttttagcccatctctgtttggcgatttagaagcg
 gatactgagacaattggatgtaatccagatgggtataacattaatgaacaatgtggctct
 actcatccagaaaaattagctgaagctgtgttagaaactgaaagtgactttggtttagct
 10 tttgatggagatggcgatcgaattattgcggtagatgaaaatggacaaattgtagatgga
 gatcaaattatgttcattatttggtcaagagatgtataaaaaccaagaactcaatggaat
 atgatagtttcgacagtaatgagtaaccttgggttctacaaagctctagaaaaagaaggt
 attcagtcaaaacaaaactaaagtggagatcgctatgttgctgaggaaatgagaagagga
 aattataatcttgggtgggtgaacaatccgggtcatatcgtaataatggattacaatactact
 15 ggtgatggattattaacgggtgttcagttggcttccgttattaaaatgagtggtaaaact
 ctaagcgagtttagcttctcaaatgaaaaagtaccacaatctttaattaatgtgagagt
 actgacaaatatcggtgtgaagagaatttcatgttcaagagataatgacgaaagttgaa
 acagagatgaatgggtgaaggaagaattcttgttcgtccttctggaactgaacctttagta
 cgtgtaatgggtgagggctgcaactgacgcggatgctgaaagatatgctcaaagtatcgct
 20 gacgttgttgaagacaaaatgggcttagataaataa

Sequence 318
 MGKYFGTDGVRGVANQELTPELAFKLGRYGGYVLAHNKGEKHPRLVGRDTRVSGEMLES
 ALIAGLISIGAEVMRLGVISTPGVAYLTKEMEAALGVMISASHNPVADNGIKFFGSDGFK
 25 LSDDQENEIEQLLDQTNPDLP RPVGEDI VHYSDYFEGAQKYL SYLKSTVDVNFEGKLIVL
 DGANGSTSSLAPFLFGDLEADTETIGCNPDGYNINEQCGSTHPEKLAEAVLETESDFGLA
 FDGDGDRII AVDENGQIVDGDQIMFIIGQEMYKNQELNGNMIVSTVMSNLGFYKALEKEG
 IQSNKTKVGDYVVEEMRRGNYNLGGEQSGHIVLMDYNTTGDGLLTGVQLASVIKMSGKT
 LSELASQMKKYPQSLINVRVTDKYRVEENIHVQEIMTKVETEMNGEGRILVRPSGTEPLV
 30 RVMVEAATDADAERYAQSIADVVEDKMGLDK*

Sequence 319
 35 Contig_0466_pos_1040_24,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P39754|GLMS_BACSU GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE A
 MINOTRANSFERASE (ISOMERIZING) (EC 2.6.1.16) (HEXOSEPHOSPHATE
 AMINOTRANSFERASE) (D-FRUCTOSE-6- PHOSPHATE AMIDOTRANSFERASE
 40) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERAS
 E) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
 atgttacaaactacaaaccaatacaaaagatacatgaccatgaaatagttattgttaag
 cgagacacagtagaaattaaagatcttgaggggcacattcaacaacgtgatacgatatacg
 gcagaaatagatgctgctgatgcagaaaaaggcgtatatgatcattacatgttaaaagaa
 45 attcatgaacagcctgcagtgatgcgtcgcatatttcaagaatatcaagatgaaaaagg
 aatttaaaaatcgattcagagattattaatgatgtagcagatgctgatcgattttacatc
 gttgcagctggtagttatcatgctggattgggtgtaagaatttattgaaaaatgg
 gcaggtgtacctactgaggttcattgtagcttctgaatttgtatataatatgccacttctt
 tctgaaaaaccactattttatttatatttcacaatctgggtgaaacagctgatagtcgtgct
 50 gtattagttgaaacaaataagtttaggtcacaatcattaacaattactaatgttgctgg
 tcaacattatcacgtgaagcggatcacattactttacatgctggacctgagattgca
 gtcgcatctacaaaagcatatacagcgcaaatgctgttttatctatcttctcaaatt
 gttgctaaaaatcatggctgctgaaaccgatgttgatttattaagagaactagctaaggt
 actacagctattgaaacaattgttgacgatgcacctaagatggagcaaatgcaacggat
 55 ttcttaaaaactctgtaatgcattcttcattggacgacaattgattataatgttagt
 ttgaaggtgcattaaaaattaaaagaaatttcttatattcaagctgaaggatttgcagg
 ggggaattaaagcacggaacaatcgctttgattgaagatggcacacctgttataggttta
 gctacacaagaaaacgttaattctatcaattcgtggaatatgaaagaagtacttttag

Sequence 320

MLQTTNQYKEIHDHEIVIVKRDTVEIKDLEGHIQQRDTYTAEIDAADAEGVVDHYMLKE
 IHEQPAVMRRRIQEQDEKGNLKIDSEIINDVADADRIYIVAAGTSYHAGLVGKEFIEKW
 AGVPTEVHVASEFVYNMPLLSEKPLFIYISQSGETADSRVAVLVETNKLGHKSLTITNVAG
 5 STLSREADHTLLLHAGPEIAVASTKAYTAQIAVLSILSQIVAKNHGRETDVDLLRELAKV
 TTAIETIVDDAPKMEQIATDFLKTRNAFFIGRTIDYNVLEGALKLKEISYIQAEFGAG
 GELKHGTIALIEDGTPVIGLATQENVNLSIRGNMKEVL*

Sequence 321

10 Contig_0467_pos_8435_9724,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P30949|GSA_BACSU GLUTAMATE-1-SEMIALDEHYDE 2,1-AMINOMU
 TASE (EC 5.4.3.8) (GSA) (GLUTAMATE-1-SEMIALDEHYDE AMINOTRANS
 FERASE) (GSA-AT). >pir:pir|D42728|D42728 glutamate-1-semiald
 15 ehde 2,1-aminomutase (EC 5.4.3.8) - Bacillus subtilis >gp:g
 p|M57676|BACHEMAXC_6 Bacillus subtilis hemAXCDBL gene cluste
 r. NID: g143034. >gp:gp|Z99118|BSUB0015_77 Bacillus subtilis
 complete genome (section 15 of 21): from 2795131 to 3013540
 . NID: g2635200.
 20 atgaaatttactgaaagcgaacgtcttcagcaactttctaataatataatgttgggaggt
 gtgaattcaccttcaaggtcgtataaagcagtaggtggtggggcaccagtagtaatgaa
 gaaggtcattgagcgttcttatgatgatggttaataataatataatcgactatctaca
 gcttatggtccaataataactggtcatgcaccccacatatcaccgaagctatccaagat
 caagcagcaaaaggcgtactttatggtaccctactgaattagaaataaatttctcaaaa
 25 aaacttagagaagcagttccttcttagaaaagattcgttctcgtaactctggtactgaa
 gcagttatgacaacaattagagttgctcgtgcttatactaaaagaaacaaatcattaag
 tttgcaggtccttatcatggtcattctgatttagtttagtggcagctggaagtggacct
 tctcaacttgggttctccagattctgctggtgtcccccaggtgttcacaaagaggttatt
 acagtaccgttttaataatgatatagaatcatatagagaagctattgattattggaagacgac
 30 attgctgcagttatttagtagagccgattgtgggtaattttgggatggtcatgccacaacca
 ggtttcttagaagaagtaataaaaatttctcatgataatggaacattagttatctatgat
 gaagttatcactgcttttctggttccattatggtgcagctcaagatttattaggtgttaaa
 ccagacctcactgcttttggtaagattgttggcgggtggtttaccaattggaggctatggt
 ggtcgacaagatattatggagcaggttgcaccattaggtccagcttatcaagcaggaaca
 35 atggccggttaaccgttatctatgagagcaggtattgctttattagaggtacttgaacaa
 gaaggtgtttatgataaaacttgatcaattaggtcgtcgtcttgaagaagggttacaaaa
 ttaatagataagcatcatattacagcaacaataaatcgaatctatgggtcactgacattg
 tatttcacaaatgaaaaagttacacattatgaacaagttgaaaactctgatggagatgct
 ttcgctcaattctttaaatatgttgaaccaaggcattaatctcgcgcttctaaattt
 40 gaagcatggttcttaactacagaacatactgaagaagatatcgatcgcacactagaagca
 gctgattatgcatttagtaaaatgaaataa

Sequence 322

45 MKFTESERLQQLSNEYILGGVNSPSRSYKAVGGGAPVVMKEGHGAYLYDVGDKYIDYLQ
 AYGPITGHAPHITEAIQDQAAGVLYGTPTELEINFSKKLREAVPSLEKIRFVNSGTE
 AVMTTIRVARAYTKRNKIKFAGSYHGHSDDLVLVAAGSGPSQLGSPDSAGVPQSVQAEVI
 TVPFNDIESYREADIYWKDDIAAVLVEPIVGNFGMVMPQPGFLEEVDNKISHDNGTLVIYD
 EVITAFRFHYGAAQDLLGVKPDLTAFGKIVGGGLPIGGYGGGRQDIMEHVAPLGPAYQAGT
 MAGNPLSMRAGIALLEVLEQEGVYDKLDQLGRRLEEGQLKIDKHITATINRIYGSLLT
 50 YFTNEKVTHYEQVENSOGDAFAQFFKLMLNQGINLAPSKFEAWFLTTEHTEEDIDRTLEA
 ADYAFSKMK*

Sequence 323

55 Contig_0467_pos_10082_11125,
 putative peptide of unknown function
 atgtctatcgcttcttcttacttctgataatattggtctaaaaacgtttagcaggtgtcag
 gcagtagttgcatgcaaccgaggtgtttatcgctcaatcaaaactgtttctgaacaagct
 attggtaatgtgattggtgcattacttgcagtaacaatggtaacgatattcaataataat
 ttcattatcatggcgcttaccgttattttactcattgcaattttgttccaatttaattctt

gcccatgtagcaacacttgcaagcgtaactgcacttataattatggggcaacacactggt
 tctttctatgttgctgcatttttttagatttgtagtgattggtgtattgagttct
 tctgttgctcaatctaatttttttacctcctaagtttgaaacaaaaatttattataattct
 atgaatatttcttctgatataatttgttgggttaaaacttgtagtcaatgacacatcagaa
 5 tttcataatatataacaggatggtgatcaactaaactcacgcacataaattagaaaag
 attttcgactattacaatgaagaaagaccattaacaaaaaacatatttatcaacagaat
 agaaaaaaaatactatttagagaagtagttagaacgaccagacaagcatatgaagtgcta
 aaacgaatgtcacgatatacaaatgatttatatcaactaaataatcaattacttttaca
 atcaaaattagaacttgattcattagttactttacatgaacaaatatttaagagtctatca
 10 aaaaaagctagatatgatgtcactcaattagattatgaagttgacaatcctcagaagaaa
 aacttgatggatgcttttcagcaagaattaattaaaaaccacatcagacgcaatattct
 tatagcaatatgatgcaaattattgctgaaattgaagaatacagatatcaacttgaacac
 ttagatagaatccgtttaagttttacctatcaccgttctgatactgatatagacatt
 tcaaatgaggactttgacttataa

15

Sequence 324

MSIASLLPDNIGLKTLAGVSAVVAMQPSVYRSIKTVSEQAIGNVIGALLAVTMVTIFNNN
 FIIMGVTVILLIAILFQFNLAHVATLASVTALIIMGHTGSFYVVAFFRVLVMIGVLSS
 SVVNLIFLPPKFETKIYYNSMNISSDIFVWFKLVNDTSEFHNKIQDGDQLNSRINKLEK
 20 IFDYNEERPLTKKHIYQONRKILFREVVRTTRQAYEVLKMSRYQNDLYQLNNQLLLQ
 IKLELDSLVLTHEQIFKSLSKKARYDVTQLDYEVDNPPQKNLMDAFOQELIKNPHTQYS
 YSNMMQIIAEIEEYRYQLEHLDRILSFFTYHRSOTDIDISNEFDL*

Sequence 325

25 Contig_0467_pos_12931_11285,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P45861|YWJA_BACSU HYPOTHETICAL ABC TRANSPORTER ATP-BI
 NDING PROTEIN IN ACDA 5'REGION. >pir:pir|S55415|S55415 ABC t
 ransporter - Bacillus subtilis >gp:gp|Z49782|BSDNA320D_2 B.s
 30 ubtilis chromosomal DNA (region 320-321 degrees). NID: g8537
 52. >gp:gp|Z99123|BSUB0020_20 Bacillus subtilis complete gen
 ome (section 20 of 21): from 3798401 to 4010550. NID: g26362
 40.
 atgctcatacctttatttgattaaatatgctatagatggcgtgattaataatcattcgctt
 35 acaaatcaagaaaaatttagtcaccttggtgtagcaataggaattgcattatttattttc
 ttaattgttcgcccgcgattgagtttatttagacaatattagctcaatggacaagtaat
 aaaatactatatgatattcgtaaacattgtataatcacttgcaagcactaagtgttcgc
 ttttatgcaataatcaagtcggtcaagtcatttcaagagtgattaatgatgtcgaacaa
 acaaaagactttattcttactggattgatgaatatctggcttgactgtataacgattatt
 40 atcgcaactttctattatgttcttccttgatgtaaaattgacgtttgctgcaattttatt
 tttccattttataattttaactgtttattttttcttggagattacgaaaacttacacgt
 gtgcgctcacaagctctagcagaagtacaaggtttcttacctgagcggttcaaggaatg
 tctgttattaaaaagttttgctattgaagacaatgaagctaaaaattttgataaccataac
 aagaattttttacaacgagccttccaacatacaagatggaacgcataattcttttgctgct
 45 attaatactgtttacagatttaggcccaataattgtgattggcgtgggttcatatttgga
 attacaggatcgattactgtcggaactctagcagcatttgcggttatctagaacaatta
 tttggaccacttagaagactagtatcttcatttactacacttacacaaagttttgcatct
 atggacagagtatttcagttaatggatgaggattacgacatcaaaaatggcattggagca
 cagccaattaaaaatcagtaagggtcaaatgtatttaaaacatgtgagtttcaaatataat
 50 gaaaaatgaaaaagaagtattacacgatattaatttaacaattaacaaaggcgaaactgta
 gcatttgtaggtatgagtggtgggaaaatctactttgattaatcttataccaagattt
 tatgatgttactcaagggtgaaatacttatcgatcatcataatgttaagatttcttaact
 ggtagtttaaggaatcaaataggcttagtacaacaagataatattctttttctgatacg
 gttaaggagaatattttgttgggtaggcctgatgagcactgatgatgaagtcgtagaagct
 55 gcaaaaatggcgaatgcccatgattttatttcaaatttaccgaatggatatgatactgaa
 gtaggagaacgaggagttaaatattctggtgggacaaaaacaaagggttgtaattgcacgt
 atcttttttaaaatcctcctgttttaatattagatgaagcaacaagtgcatggattta
 gagagtgaagctattattcaagaagcacttgatgttttaagtaaggatagaacaacatta
 attgttgcacatcgtctatctaccattactcatgcagatagaatagttgtaattgaaaaat

ggacgaattgttgagactggcacacaccaacaattaattaataaacgtgggtgcttatgag
catctttatagatttcaaaatttataa

Sequence 326

5 MLIPLLIKY AIDGVINNHS LNQEKFSHLGVAIGIALFIFLIVRPPIEFIRQYLAQWTSN
KILYDIRKQLYNHLQALS VRFYANNQVGVISRVINDVEQTKDFILTGLMNIWLCITII
IALSIMFFLDVKLTFAAIFIFPFYILT VYFFFGRRLRKLTRVRSQALAEVQGFLHERVQGM
SVIKSFAIEDNEAKNFDNHNKNFLQRA FQHTRWNAYSFAAINTVTDLGPIIVIGVGSYLA
ITGSITVGT LAAFVGYLEQLFGPLRRLVSSFTT LTQS FASMDRVFQLMDEYDIKNGIGA
10 QPIKISKQID LKHVSFKYNEKEVLHDINLTINKGETVAFVGMSSGGKSTLINLIPRF
YDVTQGEILIDHHNVKDFLTGSLRNQIGLVQQDNILFSDTVKENILLGRPDATDDEVVEA
AKMANAHDFISNLPNGYDTEVGERGVKLSGGQKQRLSIARIFLNNPPVLILDEATSALDL
ESEAIQEA LDVLSKDR T TLIVAHRLSTITHADRIVVMENGRIVETGTHQQLINKRGAYE
HLYSIQNL*

15

Sequence 327

Contig_0467_pos_6847_6395,

is similar to (with p-value 6.0e-58)

>sp:sp|P71086|FUR3_BACSU FERRIC UPTAKE REGULATION PROTEIN HO
20 MOLOG 3. >gp:gp|Z99108|BSUB0005_141 Bacillus subtilis comple
te genome (section 5 of 21): from 802821 to 1011250. NID: g2
633055. >gp:gp|Z82044|BSZ82044_9 B. subtilis 25 kb genomic DN
A segment (from sspE to katA). NID: g1673387.
gtgagtgcggaacttgaatctattgatcatgaacttgaagagtcaattgcttcattaaga
25 aaagcgggcttcgcattacaccccaaagacaagcaattatgcgttatcttatatcttca
cattcacatccaacagcagatgaaatatatcaagcactttcacctaaatttcctaata
agtgttgctactatctataataatctaagagtttttaagatatgtgtatagtaaagag
ttaacatatggtgattcatctagtaggtttgattttaatacacataatcactaccatatt
atatgtgaaaaatgtggtaaaatcggttgacttccattatccacaattagatgaagtagag
30 caattagctcaacatgtaacagattttgatgttactcatcatcggtggaaatatatgga
gtatgtaaagaatgtaaagaagaaggaaattga

Sequence 328

35 VSAELESIDHEEESIASLRKAGVRITPQRQAIMRYLISSHSHTADEIYQALSPKFPNI
SVATIYNLNRVFKDIGIVKELTYGDS SRFDFNTHNHYHIICEKCGKIVDFHYPQLDEVE
QLAQHVTD FVDVTHRMEIYGVCKECKEEGN*

Sequence 329

Contig_0467_pos_4571_4227,

40 putative peptide of unknown function

gtgacaaaccggaggaaggtggggatgacgtcaaatacatcatgcccttatgatttgggc
tacacacgtgtacaatggacaatacaaaagggtagcgaaccgcgaggtcaagcaaatcc
cataaaagttgttctcagttcggttagtctgcaactcgactatatgaagctggaatcg
ctagtaatcgtagatcagcatgctacgggtgaatacgttcccggtcttgtacacaccgcc
45 cgtcacaccacgagagtttgtaacacccgaagccggtggagtaaccatttggagctagcc
gtcgaaggtgggacaaatgattgggtggaagtcgtaacaaggtag

Sequence 330

50 VTNRRKVGMTSNHHPYDLGYTRATMDNTKGSETARSSKSHKVVLSDDCSLQLDYMKLES
LVIVDQHATVNTFPLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVVTR*

Sequence 331

Contig_0468_pos_6704_7495,

is similar to (with p-value 2.0e-82)

55 >sp:sp|O33812|YLAC_STAXY HYPOTHETICAL TRANSCRIPTIONAL REGULA
TOR IN LACR 5'REGION (FRAGMENT). >gp:gp|Y14599|SXLACRPH_1 St
aphylococcus xylosus lacR, lacP, lacH genes and 2 ORF's. NID
: g2462702.
atgaaaacagaaagttttactagagcagcagaaaatttatatacttcgcagccttctgtg

agtcgtgatattaaacggttagaattaaaatataatgttaaaatatttgaatttaaattct
 ccatattttaaactaactagagatggcgaaaagctattacaatacgcattgcaacgggaa
 agtattgaacaagaattatggcaaaacttaacatcggaatctgaaatcatctcaggcacc
 ttaacaattggaagcagttatacatatggtgaatatttattatcagaacagcttaccagt
 5 cttatgcaacaataccctaagttacatattcatttacgtgttaataattcagattctgtt
 ataaatgatattaaacacacagagtagatataggtattgtagaaaaggaaattcaagac
 aatgcaataaaaatgtaaggaaaataatggaagacgaaatggtgtatatttcaaaaaatcg
 attcaacctagaaatggatatatgtttcgtagagaaaaagggctctggaacaagggtttat
 caggaagtaggtctttctgagttgaaattaaatccatattttagatagaaattaacaatatt
 10 aagattattaaacaaatggtagaggctggaaatgggtttgcaattatttcaaaatcagca
 cttcatccagaagattatgaaaaattaatgataacaactttaaatgtgaaacgtcactat
 taccttgctcaacatgttgataaatataggtgaaaatattagagctgtcattgaaatg
 attatgaagtag

15 Sequence 332

MKTESFTRAENLYTSQPSVSRDIKRLELKNVKKIFEKSPYLKLTRDGEKLLQYALQRE
 SIEQELWQNLTSSEIIISGTLTIGSSYTYGEYLLSEQLTSLMQQYPKLHIHLRVNNSDSV
 INDIKHNVRVDIGIVEKEIQDNAIKCKEIMEDEMVIYKKSIQPRMDICFVREKSGSTRFY
 QEVGLSELKLNPLYLIEINNIKIIKQMV EAGNGFAIISKSLHPEDYEKLMITTLNVKRHY
 20 YLAQHVDKYIGENIRAVIEMIMK*

Sequence 333

25 Contig_0468_pos_14619_13816,
 is similar to (with p-value 8.0e-99)
 >gp:gp|U92974|LLU92974_13 Lactococcus lactis unknown gene, p
 artial cds, and HisC (hisC), unknown, HisG (hisG), unknown,
 HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF),
 30 HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB),
 LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB),
 IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR)
 genes, complete cds. NID: g2565137.
 atggattatagagtactactttattataaatatgtaactatagatgacctgaaactttt
 35 gcagccgaacattttgaaattttgtaagggaacatcattttaaaggagaatactagtttca
 acggaaggcattaatggaacattatctggaacaaaagaagatactgataaatatagag
 catatgcatgcagatagtcgttttgctgatttaactttttaaattgatgaagctgaaagt
 catgctgtttaaagaatgcacgtgcgtccaagacgtgaaattggtgcacttgacttagaa
 gaagatattaatccacgtgaaattaccggtaaatactattctcctaaagaatttaaagcc
 40 gcactagaagatgaaaatactgttatattagatgctcgaaatgattatgaatacagattta
 ggacatttccgtggagctattcgtcctgatataacacagattccgtgacttacctgaatgg
 gtgcgtaataataaaagaacaactcgacggaaaaaatattgtcacatattgtacaggtggc
 attcgttgtgaaaaattttctggttggttagtaaaagaaggatttgaaaacgtaggctcag
 ttgcatggtggtattgtacatacggtaaaagaccctgaaactaaagggtatattgggat
 45 ggtaagatgtatgtatttgatgaacgtattagtgtcgatgtgaatcaaattgataaaaca
 gtcacggaagagcattttgatggtactaaatactgtcttattctaaacctagtatat
 cagtattttttacaatcgttctaa

Sequence 334

50 MDYRVLLYYKYVTIDDPETFAAEHLKFCKEHHLKGRILVSTEGINGTLSGTKEDTDKYIE
 HMHADSFRADLTFKIDEASHAFKKMHVRPRREIVALDLEEDINPREITGKYYSPEFKA
 ALEDENTVILDARNDYEYDLGHFRGAIRPDITFRDLPEWVRNNKEQLDGKNIVTYCTGG
 IRCEKFSGWLKVEGFENVGQLHGGIATYGKDPETKGLYWDGKMYVFDERISVDVNQIDKT
 VIGKEHFDGTYCLILNLVYQYFLQSF*

Sequence 335

55 Contig_0468_pos_4208_3876,
 putative peptide of unknown function
 gtgtagcagaggttaagttgtcgtctaaaaacaaaaagcgatagcgttagccattggt

gcactgtcagtagataaatctattcgaaccaaggttaggtcaagccctgtttaaagct
gtagaagaacgtgctaaagaacaaggctattgtgctatttttgtaataatcatcctcag
tactttgagaaatctgattatgaagcagccatttatataatatacatatagaagaaaa
cgaaatcatcaatcattattagtaaaatttctaaaaccagttcaaaatgaatggctgga
5 atgacggtgtattatccggaagtactggattga

Sequence 336

VLAEVKLSSKNKKAIALAIGALSVDKSIRNQGLGQALLKAVEERAKEQGYCAIFVNNHPQ
YFEKSDYEAAHLYNIHIEEKRNHQSLLVKFLKPVQNEWSGMTVYYPEVLD*

10

Sequence 337

Contig_0468_pos_3149_2094,

is similar to (with p-value 1.0e-34)

>sp:sp|P09122|DP3X_BACSU DNA POLYMERASE III SUBUNITS GAMMA A
15 ND TAU (EC 2.7.7.7).

atggatcaagcaatagcgtttggagacgaacgacttactttacaagatgctttaaatggt
acaggtagtggtgatgaagcggcattaaatgagttatttaatgacattgtaaaaagtgat
gttaaagccgcatttaatagatatcatcattttatttcagaaggtaaaagaagtcaacaga
ctcattaatgatatgatttactttgttagagatacaattatgaataaaacgtctaacgaa
20 tccgttcattttgaatcacttattcatttcgacttagatatgttatacaggatgatagat
atcatcaatgatacactagtatccattaggttcagtgtaaatcaaagtgttcattttgaa
gtgttgctagttaaacttgcagaaatgattaagacacagcctcaaactgtacaaaatgta
gcaacagcatcggtagctaatagaaccagataatgagatgttattacaacgtttagaacaa
cttgaatgatgcttaaaaaccttaaaagaacaaggatcaaaactaataaagttagtcaa
25 caacctaaagaaaccaacacgtacgattcaacgatctaaaaatacgttttctatgcaacaa
atagcgaaagtattagacaaagcaacaaagatgatatacaattgttgaaagaaccattgg
caagaagtgattgatcatgcaaaaagttaataaaaaagtccttagtaagtttgctactg
aattcagaaccagtagcagctagtgaaatcatgtgttagttaaatgtgaagaat
cattgtgaaatagtaataaagatgatgaaaagagaaacaatattgaaagtgtagttgt
30 aatagagttataaaaactgtcaaagttagttggagtgcggctgaccaatggctgagagt
agagcagtagtacttacaatacgttaacaccaatgaaacacatcaaagcgaacaaacaaagc
acacaacagtcacacaaatagatattgtctaaaaagccaaaatattgaagtgtttacga
cacgaccagatacaatctatggtacttctttcttag

35 Sequence 338

MDQAIAFGDERLTQLDALNVTGSVDEAALNELFNDIVKSDVKAFFNRYHHFISEGKEVNR
LINDMIYFVRDPTMNKTSNESVHFESLIHFDLDMLYRMIDIINDTLVSIREFSVNQSVHFE
VLLVKLAEMIKTPQPTVQNVATASVANEPPDNEMLLQRLQLENELKTLKEQGIKTNKVSQ
QPKKPRTIQRSKNTFSMQQIAKVLDKANKDDIKLLKNHWQEVIDHAKSNDKKSLSVLLL
40 NSEPVASEDHVLVKFDEEIHCEIVNKDEKRNNIESVVCNIVNKTIVKVVGPADQWLRV
RAEYLQNRNTNETHQSEKQSTQQSQIDIAQKAKILKCLRHDQIQSMVLLS*

Sequence 339

Contig_0468_pos_2070_451,

45 is similar to (with p-value 0.0e+00)

>sp:sp|P36430|SYL_BACSU LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4)
(LEUCINE--TRNA LIGASE) (LEURS). >pir:pir|A41882|A41882 leuci
ne--trna ligase (EC 6.1.1.4) - Bacillus subtilis >gp:gp|M885
81|BACLEUS_1 Bacillus subtilis leucyl-transfer RNA synthase
50 (leuS) gene, complete cds. NID: g143147.

gtgaatgaaattacgacaagtataaagaacaagaagtcataattgtatcaaaatgaagca
tcaaaaaaatctgatttagaacgtacggacttagctaaagaaaaaacaggtgtgtttact
ggaacatttgcataatccgctctctgagcgtataaattactatttggatagcagattat
gttttatcaacttacgggtactgggtgcagtaattggctgtgcctggacatgatgagcgagat
55 catgaatttgcacgaagtttaatttaccattatcgaagttatagagggtggcgaagtt
caaaaatatacacaacaggtgaaggaaaacacattaattctggagaatttagacgggtcta
gaaaatgaagcggcaataagtaaacgatagaattgcttgaatctaaaggtgctgggtgag
aaaaaagtcaattataaattacgtgattggttatttagtaggcaacgttattggggagag
ccaattcctattatacattgggaagatggatcaatgactacagttcctgaagatgaattg

cctttactacttcctgaaacagatgaaattaagccatcagggtaccggtgaatctccactt
 gcaaatatagatgcgttcgtaaactgttatcgatgaaaagacagggtatgaagggcgccga
 gaaaccaataacaatgcctcaatgggctggcaggttgctggtactatttacgttacattgat
 5 ccacataacgaaaaaatgatagcagatcctgaaaaattaaagcattggctacctgttgat
 ttatatattggaggcgtggaacatgcagttacttacttattatatgcaagattctggcat
 aaagtgttatatgacttaggtgtgtaccaacaaaagaaccattccaaaaactatacaat
 caggggaatgatttttaggcgaaggcaatgaaaaatgagtaagtctaaaggtaatgtgatt
 aatccagatgatattgttgcatcacatgggtgctgatacattacgactatatgaaatgttt
 10 atgggaccttttagatgctgcatcgcatggagtgaaggggttagatggttctagaaga
 ttcttagatcggtttggagacttatcattactgatgaaaattcaatcaataaaaaaatt
 gtgatttctaacaatcattcacttgataagggtttacaactcaaaactgtgaaaaagtaaca
 gaagattttgatacacttagttttaataactgcaatcagtcattaatggtgtttattaat
 gagtgttataaaaactaatgaagtttacaaccttatatcgaaggggtttgtaaaaatggtta
 tcgcttattgcaccacacattgggtgaagaattatgggacgattaggggcatgaaaatacc
 15 attacttatcaaccatggccaacatttgatgaaagtttatttagtagatgatgaagttgaa
 atcgtagttcaagtcaatggtaagtttagagcaaaaatcaatattccaaaagatttatct
 aaagaagaatgcaagacttagccttgtctaataatgtaaaatgagttatgaagga
 aaagaagttaaaaaagttattgctgtacctcaaaagctagttaatatagttgctaaataa

20

Sequence 340

VNEITTS DKEQEVKLYQNEASKKSDLERTDLAKEKTGVFTGTFAINPLSGDKLPWIADY
 VLSTYGTGAVMAVPGHDERDHEFATKFNLP IIEVIEGGEVQKYAYTGEGKHINS GELDGL
 ENEAAISKAIELLESKGAGEKKVNYKLRDWLFSRQRYWGEPIPIIHWEDGSMTTVPEDEL
 25 PLLLPETDEIKPSGTGESPLANIDAFVNVIDEKTMKGRRETNTMPQWAGSCWYYLRYID
 PHNEKMIADPEKLKHWLPVDLYIGGVEHAVLHLLYARFWHKVLYDLGVVPTKEPFQKLYN
 QGMILGEGNEKMSKSGNVINPDDIVASHGADTLRLYEMFMGPLDAAIAWSEKGLDGSRR
 FLDRVWRLIITDENSINKKIVDSNNHSLDKVYNQTVKKVTEDFDTLSFNTAISQLMVFIN
 ECYKTNEVYKPYIEGFVKMLSPIAPHIGEELWDRLGHEINTITYQPWPTFDESLLVDDEVE
 30 IVVQVNGKVRKINIPKDLKSKEEMQDLALSNDNVKMSIEGKEVKKVIAVPQKLVNIVAK*

Sequence 341

Contig_0469_pos_1346_4051,
 35 is similar to (with p-value 0.0e+00)
 >sp:sp|P09339|ACON_BACSU ACONITATE HYDRATASE (EC 4.2.1.3) (C
 ITRATE HYDRO-LYASE) (ACONITASE).
 atggcttctaatattaaagaacaagcaaaagaacaattcgaattaaatggccaatcatat
 acttactatgacttacaacatttagaagaaaaaggctagctaaaatttctaaattacca
 40 tactcaattcgcgattgttagaatctgtgttacgacaagaggatgattttgttataaca
 gatgatcatatcaaaagcattaaagtaaaattcggaattgcaggtaacgaaggtgaagttcca
 ttcaaaccttctagagttattttacaagactttacaggtgtgccagcagtagtagatttg
 gcttctttacgtaaaagctatgaatgatgttggtggagatattaataaaatcaaccagaa
 gtacctgtggatttagttatcgaccattcagttcaagttgatagttacgctaataccagaa
 45 gcattagaacgtaatatgaaattagaatttgaacgtaactatgaacgttatcaattttta
 aactgggcaacaaaagcttttgataactataatgcagttacctcctgctacaggtattgtc
 catcaagtaaaacttagagtatattagcaaatgtagtacatgtaagagatgttgatggtgaa
 aaaacagcatttctgaacttttagtaggtactgattcacatactacaatgattaatggt
 attggtgttctaggttggggcgttgggtggtatcgaagccgaagcaggtatgttaggacaa
 50 ccatcatatttcccaattcctgaagttatcgagtgctgttaactcactctttaccacaa
 ggctcaacagctacggatttagctttacgtgtgactgaagaattacgtaaaaaagggtgta
 gttggaaaatttgtgaattcttcggtccaggtgttcaacatttaccatttagcagacaga
 gctacaattgctaacaatggctccagaatatggtgcaacgtgtggtttcttcccagtagat
 gaagaatcattgaaatatatgaaacttacaggccgtgacgaagaacatattgaattggtt
 55 aaagaatatattacacaaaaccatatgttctttgatgtagaaaaagaggatcctgaatat
 acagatgttattgatttagacttatctacagtagaggcatcactttctgggtccaaagcgt
 ccacaagacttaattttcttaagtatatgaaaaagaatttgaaaaatcagtaactgct
 cctgctggtaatacaaggacatggacttgatcaaagtgaatttgataaaaaagcagaaatt
 aattttaatgatggatctaaagcaacaatgaaaacaggagatatagcaattgctgctatt

acctcatgtactaacacttctaataccatatgttatgttaggtgctgggttagttgctaaa
aaagctgtagaaaaaggattgaaagtaccagagtttggttaagacgtcacttgctccaggt
tcaaaagttgttacaggatatattaagagattctggattacaacagtagtttagatgattta
5 ggtttcaatcttggttggttatggttgactacatgtattggttaactcagggccactatta
cctgaaattgaaaaggcagttgcggatgaagatttattagtaacttcagttttatcaggt
aatcgtaattttgagggcggaatccatccattagtgaagcaaaactatttagcctcacca
caacttggtgtagcttatgcgcttgctggtacagtagatattgatttacaaaatgaacca
attggtaaaaggtaaagatggtaaagatgtatattacaagacatttggccttcaatacaa
10 gaagtttctgatactgtagataaaagttgttacacctgaactattcttagaagaatataaa
aatgtatatcataacaatgaaatgtggaatgaaatagatgaaccgatgaaccattatat
gatttcgatccataattcaacatatattcaaaatccaacattttccaaggattatctaaa
gagccgggtaaaattgaaccacttaaaagtttgagagttatgggtaaaatttggtgattct
gttacaacagaccatatattctccagcaggtgctatcggtaaagatacaccagcagaaaa
tacttattagatcatgatgttgcaattcgcaactttaactcttaggttcccgtcgcggt
15 aaccacgaagttatggtacgtggtacatttgccaatattcgtagtaaaaaaccaacttgct
ccaggtactgaaggcggtattacaacatatggcctaccggagaaataatgcctatatat
gatgcagcaatgaaatataaagaagatggaactggcttagttgtcttagctggtaatgac
tatggaatgggattctctcgtagctgggctgaaaaggtagcaatttattaggagttaaa
actgtcaattgcacaaagctatgaacgtattcatcgcttaacttagttatgatgggtgta
20 ctaccgcttcaattccaacaaggagaatctgcagaagcactgggtcttgatggaaaagaa
gaaatatctgtagatattaatgaagatgtacagccacatgatcttgtaaatgtgactgca
aaaaaagaaaatgggtgaaatcattaattcaaaagctattgtacgttttgattcactagta
gaattagattattatcgtagtggtggtattttacaaatggtactaagaaataaacttgcg
cagtaa

25 Sequence 342
MASNIKEQAKKQFELNGQSYTYDYDLQTLLEEKGLAKISKLPYSIRVLLESVLRQEDDFVIT
DDHIKALSKFGNAGNEGEVFPKPSRVILQDFTGVPVAVDLASLRKAMNDVGGDINKINPE
VPVDLVIDHSVQVDSYANPEALERNMKLEFERNYERYQFLNWATKAFDNYNAVPPATGIV
30 HQVNLEYLANVVHVRDVGDEKTAFPDPLVGTDSHTTMINGIGVLGWGVGGIEAEAGMLGQ
PSYFPIPEVIGVRLTHSLPQGSTATDLALRVTEELRKKGVVGKFVEFFGPGVQHLPLADR
ATIANMAPEYGATCGFFPVDEESLKYMKLTGRDEEHIELVKEYLQQNHMFFDVEKEDPEY
TDVIDLDLSTVEASLSGPKRPQDLIFLSDMKKEFEKSVTAPAGNQGHGLDQSEFDKKAIEI
NFNDGSKATMKTGDIATAAITSCTNTSNPYVMLGAGLVAKKAVEKGLKVPEFVKTS LAPG
35 SKVVTGYLRDGLQYLDLGFNLVGYGCTTCIGNSGPLLPEIEKAVADEDLLVTSVLSG
NRNFEGRIRHPLVKANYLASPQLVVAYALAGTVDIDLQNEPIGKKGDKDGYLQDIWPSIQ
EVSDTVDKVVTPELFLEEKVNYHNNEMWNEIDVTDEPLYDFDPNSTYIQNPTFFQGLSK
EPGKIEPLKSLRVMGKFGDSVTTDHISPAGAIGKDTAGKYLLDHDVAIRNFNSYGSRRG
NHEVMVRGTANIRIKNQLAPGTEGGFTTYWPTGEIMPIYDAAMKYKEDGTGLVVLAGND
40 YGMGSSRDWAAKGTNLGVKTVIAQSYERIHRNLVMMGVLPLOFQOGESAEALGLDGKE
EISVDINEDVQPHDLVNVTAKKENGEIINFKAIVRFDLSVELDYRHHGILQMVLRNKLA
Q*

45 Sequence 343
Contig_0469_pos_4174_4641,
putative peptide of unknown function
atgatatacagtttgactgaaattgaagcaagatatcaagaaaccgataaaatgggggtt
atctatcacggtaactacgcaacatggtttgaagttgcgagaacagactatataagaaag
cttggttcagttatgcctctatggaagaacaaggtgttatttcaccagttgtagattta
50 aaagtgcaatataaaaaatcaatttactatcctgaaaaggtagacagtaaaaacatgggtg
gaaaaatattctagattacgttcaacttattgttatgaggtttataatgaaaatggagag
ttagctactactggttcaacagaacttatctgtattaaagcagatacatttaaacccata
cgcttgatagatattttcctgagtgcatgagacttatagtaaagtttaaccagttaaat
aaagaaggtaaaagatgctgaggttacgtttggcattaatcatttataa

55 Sequence 344
MIYSLTEIEARYQETDKMGVIYHGNATWFEVARTDYIRKLGFSYASMEEQGVISPVDL
KVQYKKSIYYPEKVTVKTWVEKYSRLRSTYCYEVYNENGELATTGSTELICIKADTFKPI
RLDRYFPEWHETYSKVNQLNKEGKDAEVTFGINHL*

Sequence 345
 Contig_0469_pos_6050_8044,
 is similar to (with p-value 0.0e+00)
 5 >sp:sp|P50072|PARE_STAAU TOPOISOMERASE IV SUBUNIT B (EC 5.99
 .1.-). >gp:gp|D67075|D67075_1 Staphylococcus aureus DNA for
 DNA topoisomerase IV GrlB subunit, DNA topoisomerase IV GrlA
 subunit, complete cds. NID: g1777319. >gp:gp|L25288|STAGYRA
 10 SL_1 Staphylococcus aureus gyrase-like protein alpha and bet
 a subunit (grlA and grlB) genes, complete cds. NID: g561878.
 >gp:gp|A48501|A48501_1 Sequence 3 from Patent WO9603516. NI
 D: g2302280.
 atgaataaacaaaataattattcagatgattcaattcaggtacttgaaggactagaagca
 gtttaggaagagacctggtatgtacattggatcaactgataaacgaggattacatcatctt
 15 gtatatgaagttgtcgataactccgtcgatgaagtattaaatggttatggtgatgcgatt
 acagtaacaattaatcaggatggttagtatttctatagaagataatggtcgaggtatgcc
 acaggtatacatgcgtctggcaaacctactgcagaagttatatttactgttttacatgct
 ggaggttaaatttggacaaggaggttataaaacatctggaggtctccatggggtgggtgct
 tctgtagtaaattgcccttagtgaatggcttgaagttgaaattcatagagatggtaatac
 20 tacacacaaaatttcaaaaatggtggtattccagcgacaggtttagtaaaaaactggaaa
 acaaaaaaaaactggtactaaagttacatttaaaccagactcagaaatatttaagtcaacg
 acgacttttaattttgatattttaagtgcgctttacaagaatctgcatttttacttaaa
 gatttaaaaaattacacttactgatttacgtagtggaagaacgagaagaaatttaccat
 tacgaagaaggaattaaagaatttgttagttatgtcaatgaaggtaagaagtattacat
 25 gatgttactacatttgcagggcattccaatggaatagaggtagacgttagcattccaatat
 aatgttcagtactctgagagcatattaaagtttgtaaataatgttcgtacaaaaggacgga
 ggtactcatgaagttggtttcaaaacggcgatgactcgtgttttaataatgaatatgcacgt
 cgtataaacgaactgaaagataaaagataaaaaatttagacggtaatgatatacgcgaagg
 ttaacagcgataatttcagtacgtataccagaagaacttcttcaatttgaagggcaaacg
 30 aaatcaaaacttggcacttcagaagcaaggagtgctgtagactctgttgttcagaaaaa
 ttaccatattacttagaagaaaaggcccaattatctaatacttagttaaaaaagcaatt
 aaagctcaacaagcacgcgagggctgctcgttaaagctagagaagatgcacgctccggaaa
 aaaaataaacgtaaaagatacattgttatcaggttaagtttaactcctgcgcaagtaaaa
 actgataaaaaacgagttatatctagttgaggggtgattcagcgggaggttctgcaaaattg
 35 ggacgcgacgtaaatccaagctattttacctcttcgtggaagggttattaatacagaa
 aaggcacggttagaggatatttttaaaaaatgaagaaattaatacgattattcactactat
 ggtgctggtgttggtagtactttaaattgaggatagtaattacaacagaattattatc
 atgacagatgctgatacggatggtgcacatatcaagtattattgcttacattttcttt
 aaatatatgaaaccacttgttcaagctggacgtgtctttattgcgttaccgcctttatac
 40 aaattagaaaaagcaaggtaagaataaaaaagttgagtagcgttggactgatgaagaa
 tttagaaaaattcaaaaagcaatttaggaaaagggttcatattacagcgttataaaggctt
 ggtgaaatgaatccagaacaattatgggaaactaccatgaatccagaaactcgacatta
 attagagttcaagttgaagatgaagttcgttcatcaaaacgtgtcactactttgatgggg
 gataaggttgcacgaagagagtggttgaaaaacacgttgaatttggtatgaagaa
 45 gatcaaaagcattttggataataaagaagtcctaaatactagagaatgaaaaatatattgag
 gaggaacgaattga

Sequence 346
 MNKQNNYSDDSIQVLEGLEAVRKRPGMYIGSTDKRGLHHLVYEVVDNSVDEVLNQYGD
 50 TVTINQDGSISIEDNGRGMPTGIHASGKPTAEVIFTVLHAGGKFGQGGYKTSGLLHGVGA
 SVVNALSEWLEVEIHRDGNIIYTNFKNGGIPATGLVKTKTKTKTKVTFKPDSEIFKST
 TTFNFDILSERLQESAFLLKDLKITLTLDRSGKEREIYHYEEGIKEFVSYNVNEGKEVLH
 DVTTFAGHSNGIEVDVAFQYNVQYSESILSFVNNVRTKOGGTHEVGFKTAMTRVFNEYAR
 RINELKDKDKNLGDNDIREGLTAIISVRIPEELLQFEGQTKSKLGTSEARSADVSVSEK
 55 LPYYLEEKQGLSKSLVKKAIIKAQQAREAAARKAREDARSGKKNKRKDTLLSGKLT
 PAQSKN
 TDKNELYLVEGDSAGGSAKLGRDRKFQAILPLRGKVINTEKARLEDIFKNEEINTIIHTI
 GAGVGTDFKIEDSNYNRIIIMTDADTGAHIQVLLLTFFFKYMKPLVQAGRVFIALPPLY
 KLEKKGKGNKKVEYAWTDEELENLQKQLGKGFILQRYKGLGEMNPEQLWETTMNPETRTL
 IRVQVEDEVSRSSKRVTTLMGDKVAPPREWIEKHVEFGMQEDQSILDNKEVQILENEKYIE

EETN*

Sequence 347

Contig_0469_pos_8185_10443,

- 5 is similar to (with p-value 0.0e+00)
 >sp:sp|P50073|PARC_STAAU TOPOISOMERASE IV SUBUNIT A (EC 5.99
 .1.-). >gp:gp|D67075|D67075_2 Staphylococcus aureus DNA for
 DNA topoisomerase IV GrlB subunit, DNA topoisomerase IV GrlA
 subunit, complete cds. NID: g1777319.
- 10 atgtattcaagtgggaatacgtatgataaaaatttccgtaaaagtgcgaaaactgtcggg
 gatgtaataggtcaatatcatcctcatggagactcttcagtatatgatgctatgggtgcgc
 ttaagtcaagattggaagttacgtcatgttctaattgaaatgcatggtaataatggtagt
 atcgataacgatcctccagctgctatgcgttacacagaagctaaacttagtcaattatca
 gaagaactattaagggatattaataaggaaacagtatcatttattccaaactatgatgac
 15 acaacttttgaaccaatgggtattaccagcgagatttccctaatttattaataatggatct
 acgggggatttcttcaggatatgctactgatatcccgccgcataacctcgccgaagtaata
 caaggcacattgaagtatatcgatcaacctgatattacaattaatcaactgatgaaatat
 atcaaaggcctgactttcctacaggtgggtatcattcaaggaatagaaggatataaaaaa
 gcgtatgagaccggttaaaggaaaggttgcgtgcgttcacgagtagatgaagagccttta
 20 agaagtggacgtaaacaattaattgtgactgaaattccgtatgaagtgaataaaagttagt
 ttagttaaagaattgacgaattacgtgccgataaaaaaggttgatggattgtagaagtt
 cgagatgagactgatagaactggattacgaattgcaatcgaattaaaaaaagatgcta
 atagcgaatcaatcaaaaactatttatataagaattcggatttacaatttcatataat
 aatatggttgctatttagtgaaggtcgccctaagttgatgggattacgtgaaattataga
 25 agttattttaaattcatcaaatggaagtgttacaatagaacgcgttatgacttagagcaa
 gctgaaaaacgtatgcatattgtggaaggattaatgaaagctttatctatacttgatgaa
 gttattgcatgatacgtaatcttaaaaaataaaaaagatgctaaagataatttagttgca
 gagtatgactttactgaagctcaagcagaagctattgtcatgttacagctgtatagatta
 acaataactgacattgaagctttgaaaaagaacatgaagagttagaagctttaataaaa
 30 gaattaagaaatatcttagataatcatgaggcacttttagcagtaattaaagatgaacta
 aatgaaattaaaaagaaattttaaagtggatcgactatctacaatcgaagctgaaatttcc
 gaaatcaaaattgataaagaagttatggtgcctagtgaagaagtgattttaagtttgacg
 caacatggctatataaaacgtacatctacacgtagttttaacgcaagtggtgtgactgaa
 atcggtttgaaggacggcgaccgtttattaaaacatgaaagcgtgaatactcaagatact
 35 gttcttgattttacaaataaaggtagatattgtttatacctgttcataaattagccgat
 atccggttggaagagacttggtcaacacatatcacaattgtgccaatagatgaagatgaa
 gaagtggtaaatgtatacaacgaaaaagattttaaaaaatgaagccttttatattatggct
 acaaaaaacggcatgattaagaaaaagtagtgcttcacaatttaaaactactcggttta
 aaacaaactcaaaatatgaaggttaaagacaaagatgaacttattaatgtcggttcgatta
 40 gagtctgatcagtttaattactgttctaaccataaaaggcatgtcattaacttattcaact
 aatgaattatcggtacaggttaagagcagctgggtgtaaatcaattaatcttaagat
 gaagactatgttggtatgacagaagatgtgaacgactcagattccataataatgggtaca
 caacgtggtgctatgaagcgtattgattttaatgttcttcaagaagctaaacgcgcacaa
 cgtggaattactttactaaaagaattaaagaaaaaacgcacgaattgtggcaggtgca
 45 gtagttaaagaaaatcacacgaaatatattgtattctctcaacatcatgaagaatatggt
 aatatcgatgatgtacacttatctgaacaatatataactaatggatcatttattattgatact
 gatgattttggagaagtagaaagtatgattcttagagtaa

Sequence 348

- 50 MYSSGNTYDKNFRKSAKTVGDVIGQYHPHGDSSVYDAMVRLSQDWKLRHVLIEMHGNNGS
 IDNDPPAAMRYTEAKLSQLSEELLRDINKETVSFIPNYDDTTLEPMVLPARFPNLLINGS
 TGISSGYATDIPHNLAIEVIQGTILKYIDQPDITINQLMKYIKGPDFPTGGIIQIGIEGIKK
 AYETGKGKVVVRSRVDEEPLRSGRKQLIVTEIPYEVNKSSLVKRIDELRADKKVDGIVEV
 RDETDRTGLRIAIELKKDANSESIKNYLYKNSDLQISYNFMVAISEGRPKLMGLREIIE
 55 SYLNHQIEVVTNRTRYDLEQAERMHIVEGLMKALSILDEVIALIRNSKNKKDAKDNLVA
 EYDFTEAQAEIIVMLQLYRLTNTDIEALKKEHEELEALIKELRNILDNHEALLAVIKDEL
 NEIKKKFKVDRSLSTIEAEISEIKIDKEVMVPSEEVILSLTQHGHIKRTSTRSFNASGVTE
 IGLKDGDRLLKHESVNTQDITLVFTNKGRYLFIPVHKLADIRWKELGQHSQIVPIDEDE
 EVVNVYNEKDFKNEAFYIMATKNGMIKKSSASQFKTTRFNKPLINMKVKDKDELINVVRL

ESDQLITVLTHKGMSLTYSTNELSDTGLRAAGVKSINLKDEDYVVMTEDEVNDSDSIIMVT
 QRGAMKRIDFNVLQEAQRAQRGITLLKELKKKPHRIVAGAVVKENHTKYIVFSQHHEEYG
 NIDDVHLSEQYTNGSFIIDTDDFGEVESMILE*

5 Sequence 349

Contig_0469_pos_10802_12193,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q45068|ALST_BACSU AMINO ACID CARRIER PROTEIN ALST. >g
 p:gp|273234|BC170DEGR_21 B.subtilis DNA (26.2 kb fragment; 1
 10 70 degree region). NID: gl405443. >gp:gp|Z99113|BSUB0010_105
 Bacillus subtilis complete genome (section 10 of 21): from
 1781201 to 2014980. NID: g2634090.
 atgttaccagagatgttttagagcattaactgaaaagccagaaactttaagtagtggtgag
 aagggtatattcaccatttcaagcttttgcgattagtgctgggtcaagagtaggaactgga
 15 aatattgccgggtgttgcaactgctattgttcttgggtggcccggtgcagctcttctggatg
 tggattattgctttttattgggtgcagctagtgcatattatggaagcaacgcttgctcaagtt
 tataaggtagcatgacaaagaagggtggattccgtggcgaccagcctattacataacaaaa
 gggctaaacaaaaatggcttggaattgtatttgcgtgttttaattacagttacatttgc
 20 tttgtatttaatactgttcaagcgaatacaattgctgaatcattaaatacacataacaat
 atagcccggtgaattactggaatagtagtactgcagttattacaggtattatcatctttgg
 ggtgttcgtagcatagctacactatcttcaacttattgtgcctattatggctattgtttat
 ataggtagtggtttttaatacatttttactcaatatagatcaaattgtacctatgattggc
 actattattaaaaagtgcatcgcgagttcagcaggttactgggtggtgctgtaggagctgct
 attcttcaaggtattaaacgtgggtttattctcaaacgaagctggtagggatctgcacct
 25 aatgctgctgctacatctgctgtgccccatcccggttaacaagggtttaattcaatcatta
 ggtgtattctttgacactatgcttggttgtacagctacagcaattatgattttattatat
 tctggtttgcaatttggtagtagcgccctcaagggtgtagcagttacgcaatcagcgttg
 aacgaacatttaggttcagcaggaggtattttcttaactgtagcagttacctatttgca
 ttttcatctgtttaggtaactattactatggacaatccaatattgaatttttatctaac
 30 aataagatgatattatttttttagatggtttgtagtagtacttttagtattttaggtgct
 gttgctaaaacagaaacagtttggagtagtgcgattttatgtgggtcttatggcaata
 gtaaatatcatatcaattatagggttgcgaatattgcgtttgcagtgatgaaagattat
 caaagacaggagataatcataaaaattagatactacagtcagtagtcagattaattgtctt
 aggactaatgctttttcttttccggtgataagagaggacattctcatattcaaaagggtg
 35 tttcatccatcttataaaaatcaaaaagttatcaaaaactaacttctttatagcatatctt
 tttactttctaa

Sequence 350

MLPEMFRALTEKPELSSGEKGISPFQAFSAISAGSRVGTGNIAGVATAIVLGGPGAVFWM
 40 WIIAFIGAASAFMEATLAQVYKVHDKEGGFRGGPAYIITKGLNQKWLGIWFAVLITVTF
 FVFNVTQANTIAESLNTQYNISPVITGIVLAVITGIIIFGGVRSIATLSSLIVPIMAIVY
 IGMVLIILLNIDQIVPMIGTIIKSAFGVQVQVTGGAVGAAILQGIKRGLESNEAGMGSAP
 NAAATSAVPHPVKQGLIQSLGVFFDTMLVCTATAIMILLYSGLQFGDSAPQGVAVTQSAL
 NEHLGSAGGIFLTVAVTLFAFSSVVGNYYYGQSNIEFLSNNKMILFI FRCFVLLVVFVGA
 45 VAKTETVWSTADLFMGLMAIVNIIISIIGLSNIAFAVMKDYQRQEIIKIRYYSHSQINCL
 RTNAFFFSVIRELILFKRCFHPYKINKVIKTNFFIAYLFTF*

Sequence 351

Contig_0469_pos_12403_11954,
 50 putative peptide of unknown function
 gtgtatacttatactattaaaaattagaaatgagattattatgaaaatagtagaagtaaaa
 tctaagaatggtaccaattttatgatttttagatggttaataatgaacctatagtagatgca
 gtaagatatttgaagtatctggatagtggttaagaaaagtttaataaccaagaaaacctat
 gcctatgcactaaaaaatttttttggttacttagaaagtaaaaagatatgctataaagaa
 55 gttagttttgataactttgttgattttataagatggatgaaaacaccttttgaatatgag
 aatgtcctctcttaccgaaaagaaaaagcatttagtccctaagacaattaatctgact
 atgactgtagtagtataattttttatgattatctcctgtctttgataatctttcatcactgc
 aaacgcaatatctcgacaaacctataattga

Sequence 352

VYTYTIKIRNEIIMKIVEVKSNGTNFMILDGNNEPIVDAVRYLKYLDVKKSLNTKKTY
 AYALKNFFVYLESKKICYKEVSFDNFVDFIRWMKTPFEYENVLSYHRKEKSISPKTINLT
 MTVVSNFYDYLLSLIIFHHCKRNIRQTYN*

5

Sequence 353

Contig_0470_pos_4232_4603,

10 is similar to (with p-value 1.0e-23)

>gp:gp|D85752|D85752_8 Enterococcus faecalis plasmid pPD1 ba
 cA, bacB, bacC, bacD, bacE, bacF, bacG, bacH and bacI genes,
 complete cds. NID: g2879906.

atgatagtaattactgatgcaataccaataattattccaatcatcgtaaagatattacgt
 15 cgtttgttttttaaaatggaacgaatagcaactgatatgacatttgaaaagtattcacg
 tgtcaacacctcttcctcttgacacgccccatccaaaataggataatacagatcagcttt
 ttccgccacttcacgatcatgctgtaccataataatagttgtattctgttctttgttcag
 ttttacgaaaagctccataatatcttgagatgtcttcgaatcaagagcgccagtaggttc
 atcggcaataataaacttagggtcattaataattgcccgggcaatagctacacgttggtg
 20 ctgccctcctga

Sequence 354

MIVITDAIPIIPIIVKILRRLFFKMERIATDMTFEKLFTCQHLFLLHTPIQNMDNTISF
 FRHFTIMRYHNNSCILFFVQFYEKLNILRLRIKSASRFIGNNKLRVINNCPEGNSYTL
 25 LPS*

Sequence 355

Contig_0470_pos_6112_7041,

putative peptide of unknown function

30 gtggtttatgttggtgagttaaatatattagaatgtgagggagtatttgaattgaagaaa
 ttagcagtgatagtgcgtttacaataatattagctggttggtggtccttggtgatagt
 gataataatggaagctcaacgataaatgatgatcaacaatcaggatataaaaagtaacaga
 gattcaaaatcaagtataagtagaatacaaacagaagataatcagcaggacacacaacaa
 gataccattcgaatagataactatgctcaagtttggttaactgcttttagatagttataga
 35 ggtgaaagtgccttccttttgacgatttagaaattgtacatcaaaatatttctaataaa
 gtttttagatccctatcacccagacgaatcagccaaactacctgaaggacagaattgtta
 acagcaagtgttactgcagcagggttcagtttattataaaagtaatggagatggcacaatt
 acaatatatagtgatccatcacatttccaagggagtggcgtgacgctgattactctaaa
 agagaatctcaacgcattatagatgatgctcgtacagtttaagttatacaacgctagttaa
 40 agtgaatcaataagataagtcagatgatgaggactgaattttcagttggtgataattta
 acagatgaagatgatacttctgaatctgaagatcaatcaagtagttctgatgaagcaacg
 gtgacacgaagtaattgttatcgatatagttgaagactacgaagggcatcaattagatata
 gacacatatatttcaaaagaaccagaaaaagatagcgatggtagttgggggttctcattt
 acagataaagaaggccatttagaaggatcttatattatcgataaagatggagaagtaacg
 45 aagtatgatgaagatggagagccagaataa

Sequence 356

VVYFVELNILECEGVFELKKLAVICAFTILILAGCGLGSDNNGSSTINDDQQSGYKSNR
 DSKSSISRQTEDNQDTHSNRYAQVWLTALDSYRGESDLFPDDLEIVHQNISNK
 50 VLDPYHPDESALKEGTELLTASVTAAGSVYYKSNNGDTITIYSVPSHFQGSWRDADYSK
 RESQRIIDDARTVKLYNASESEINKISQMMRTEFSVGDNLDEDDTSESEDQSSSSDEAT
 VTRSNVIDIVEDYEGHQLDITYIYKEPEKSDSGSWGFSFTDKEGHLEGSYIIDKDGEVT
 KYDEDEGEPE*

55 Sequence 357

Contig_0470_pos_10172_10480,

putative peptide of unknown function

atgatggccaacacctttaataataaccaatacatcttccatacagagcggttcggttcatt
 ccgcgtgataatagtaatgcagttaaaataatcactacagcagcaatgatataatgaca

ccaccgttacttccaaatggattagataatgatttaggttaaagaaatgcccaatggtgca
ataagacctcttaagtttagcagaaaagcctgaagcaacgaaagcaacagcaataaagtat
tctgctaaaagcgcccaaccggcaaccatccgaataattcaccaaaaagaacattaatc
catgaataa

5

Sequence 358

MMANTFNITNTFSIRAASFIPRDNNAVKIITTAAMISMTPELLPGLDNDLGKEMPNGA
IRPLKLAEKPEATKATAIKYSAKSAQPATHPNNSPKRTLIHE*

10

Sequence 359

Contig_0470_pos_11394_12080,

is similar to (with p-value 8.0e-79)

>gp:gp|D78193|BACGNTZA_30 Bacillus subtilis 36kb sequence be
tween gntZ and trnY genes encoding 34 ORFs. NID: g1064780. >

15

gp:gp|Z99124|BSUB0021_142 Bacillus subtilis complete genome
(section 21 of 21): from 3999281 to 4214814. NID: g2636442.

atggaagaacttttcagccaaatcgacagaaacattaaggatttaaaccggaatttttagtg
acacatgaacacatcgaccatattaaaggcttgggtgttttagcacgtaaatataaactt
ccgatttacgcgaatgagaatacatggaaagcgatagagaagaagatagccgcattcca

20

atggatcagaaatcttcttaatccatagaaacgaaatctcttgacgagattgatata
gaatcatttaacgtgtcacatgacgcgattgatccacaattctacatcttcacaataac
tataagaaatttacgatgataactgacactggttacgtttcagatcgatgaaaggatg

attcaaggtagtgatgtctttatgtttgaaagtaatcacgatgtcgatattgtacgatg
tgtcgctatccatggaagacgaaacaacgtattttaagtgatatgggtcacgtatccaat

25

gaagacgcgggtcttgcatgagtgatgtcattacaggttaatacgaacgtatatacctc
tctcatttgtcacaagacaataatataatgaaagacctcgacgcgatgagtggttgacaagt
ctcaacgaacacgatatcgatacagagaaagaagtattgctttgcataccgataaagca

caagccacaccgattttatacactataa

30

Sequence 360

MEELFSQIDRNIKDLNGLVTHEHIDHIKGLGLVLRKYKLPFIYANENTWKAIEKKDSRIP
MDQKFIENPYETIKSLAGFDIESFNVSHDAIDPQFYIFHNNYKFTMITDTGYVSDRMKGM
IQGSDVFMFESNHDVMDLRMCRYPWKTKQRILSDMGHVSNEAGLAMSDVITGNTKRIYL
SHLSQDNMMDLARMVSGQVLNEHDIDTEKEVLLCDTQKATPIYTL*

35

Sequence 361

Contig_0470_pos_13070_13600,

is similar to (with p-value 1.0e-48)

>gp:gp|D78193|BACGNTZA_15 Bacillus subtilis 36kb sequence be
tween gntZ and trnY genes encoding 34 ORFs. NID: g1064780. >

40

gp:gp|Z99124|BSUB0021_128 Bacillus subtilis complete genome
(section 21 of 21): from 3999281 to 4214814. NID: g2636442.

gtgtataacttgtggataactcgaataaatcatgtacttttgaggataaaatgaagatt
actatcttatcagttggaaaactaaaagaaaaatattggaagcaagccattgcagaatat

45

gaaaaaagatttagaccttacacgaaaatcgaattaatagaagtaccagatgaaaaagca
cctgaaaatatgagcgacaaagaaatagaacaagttaaagaaaaagaaggccaacgccta

ctcaataagattaactcccaatctacagtaatcacgttggaatcaaaggcaaaatggtg
tcttcagaaggactcgctaaagaactgcaaacacgcgatgacacaaggtcaaagcgacttt

acatttgtcataggtggtccaatggtttacaccaagacgtcttacaacgcagcaactac
gcactatcattcagcaacatgaccttcccatcaaatgatgcgtgtaattattgattgaa

50

caaatttatcgcgcatcctcaaatcatgagaggtgaagcgtatcataaatga

Sequence 362

VYNLWITRINHVLLEDKMKITILSVGKLKEKYWKQAIAYEYKRLGPYTKIELIEVPDEKA
PENMSDKEIEQVKEKEGQRLNKNINSQSTVITLEIKGMVSSEGLAKELQTRMTQGQSDF
TFVIGGSNGLHQDVLQRSNYALSFSNMTFPHQMMRVILIEQIYRAFKIMRGEAYHK*

55

Sequence 363

Contig_0470_pos_15541_14693,

is similar to (with p-value 3.0e-75)
 >sp:sp|P04188|STSP_STAAU GLUTAMYL ENDOPEPTIDASE PRECURSOR (E
 C 3.4.21.19) (STAPHYLOCOCCAL SERINE PROTEINASE) (V8 PROTEINA
 SE) (ENDOPROTEINASE GLU-C). >pir:pir|A26812|PRSASK glutamyl
 5 endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aure
 us >gp:gp|Y00356|SASP_1 Staphylococcus aureus V8 serine prot
 ease gene. NID: g46686.
 atgaaaaagagatttttatctatatgtacaatgacaattgcagcgtagcaactactaca
 atggtaaatacttcttatgcaaaaaccgatacagaaagccataatcattcctcacttggc
 10 acagaaaaacaaaatgttttagatattaatagttcagagtcataatatcaaaccaagtcaa
 atataaaagttaccacagtgtaattattacctaataataatagacatcaaatttttaatact
 acacaagggtcattatgatgtctgttagttttttatataaccaatagatgggtggatatatg
 agtgggttcaggtgttgtgttaggtgaaaatgaaatattaactaataaacacgttgtaaat
 ggagctaagggtgaatccaagaaatattagtgccatccttcagctaaaaatgaaaatgat
 15 tatcctaattggcaatttgtgggtcaagaaatcataccgtatcctggtaatagtgattta
 gcaatcttaagagtgacacaaacgaacataatcaacataattgggtcaagtagttaaacct
 gcaactataagtagcaatacagacactagaattaatgaaaacatcactgttactggttac
 cctggtgacaaaccattagccacaatgtgggaaagtgtaggtaaaagttgtctatatattggt
 ggcgaggaattaagatatgacctaagtactgtaggtggaactctggatctccagtattt
 20 aatggtaaaaatcaagttattggaatacattatgggtggcgtagataataataacaatagc
 agtgttttatattaatgatttcgttcaacaattcctaagaacaatatacctgatataaat
 attcagtaa

Sequence 364
 25 MKKRFSLICTMTIAALATTTMVNTSYAKTDTESHNHSSLGTENKNVLDINSSSHNIKPSQ
 NKSYPVILPNNRRHQIFNTTQGHYDAVSFIYIPIDGGYMSGSGVVVGENEILTNKHVVN
 GAKGNPRNISVHPSAKNENDYPNGKFVGQEIIPYPGNSDLAILRVSPNEHNQHIQVVKP
 ATISSNTDTRINENITVTGYPGDKPLATMWESVGKVVIYIGGEELRYDLSTVGGNSGSPVF
 NGKNQVIGIHYGGVDKNYSSVYINDFVQQFLRNNIPDINIQ*

Sequence 365
 Contig_0470_pos_14467_13679,
 is similar to (with p-value 2.0e-43)
 >gp:gp|U56999|TPU56999_1 Treponema pallidum methyl-accepting
 35 chemotaxis protein (mcp-1) gene, complete cds, and potentia
 l regulatory molecule (pfoS/R) gene, partial cds. NID: g1354
 774.
 atgagtcattcgaataaatggaaagcgattttttaacaacattttgaatgcagtagga
 gcaggggtagttattgcactgttacctaataatgccttattaggtgaattatataaattcttc
 40 aaagaaggtaaatcatgtactagaaacgatttttcagctagtaacaatcatacaatctttt
 atggctttttattataggggttcttgctgcgcaccaatttaaatttaagggtacaggtgct
 gcaattattgggtatttcagcaatgctaggttctggagctgtacactataatggacaaaca
 attgaattaaaagggaattggagatattataaatgtaatttttagtagttatattagcgtgt
 ttcatattatatttttagaggggaaattaggttccttagaaatgattattttaccggtt
 45 ttagttcctgtaattagtggttaataagggttattaacattaccttacgttcaagttatt
 acgcagtcactaggaaaattagtaaacagggtttacagaattaaatccattattaatgtct
 atattaatttgtgtaacattttctttattaatggtaactccaatctcgttagttgctata
 gcaacagcaattaaccttactggttttaggaagtgggtgctgcaaatatgggaatagttgca
 gcttggtgtaacctttttatttggatctttaagagttaattctcttgaggttaacgtggt
 50 ttactcataggtgctgctataaatgatgatttcctgtgtacttaaagcaaaagcaaacataca
 ttcaattag

Sequence 366
 55 MSHSQINGKRFFNNILNAVAGVVIALLPNALLGELLKFFKEGNHVLETIFQLVTIIQSF
 MAFIIGVLAHQFKFKGTGAIIIGISAMLGSGAVHYNGQTIELKGIGDIINVILVVILAC
 FIYMFLEGLKLSLEMIILPVLVPVISGLIGLLTLPYVQVITQSLGKLVNRFTELNPLMS
 ILICVTFSLLMVTPISLVAIATAINLTGLGSGAANMGIVAACVTFGLGSLRVNSLGVNVV
 LLIGAAMMIPVYLKQSKHTFN*

Sequence 367

Contig_0470_pos_10545_9307,

5 is similar to (with p-value 3.0e-44)

>gp:gp|AL023702|SC1C3_2 Streptomyces coelicolor cosmid 1C3.
NID: g3169026.

gtggcagggtctttagccttttactttagcagaaatggcatctacaatgccttttgcctggg
tcagcttattcatggattaatgttcttttgggtgaattatcggtgggttgcgggttgg
10 gcgcttttagcagaatactttattgctgttgccttgcgttgcctcaggcttttctgctaac
ttaagagggtcttattgcaccattgggcatttctttacctaatacattatctaataccattt
ggaagtaacgggtggtgcatgatatcatgctgctgtagtgattattttaactgcatta
ctattatcacgcggaatgaacgaagccgctcgtaggaaaatgtattgggtatattaaag
gtgttggccatcattttatttgtgattgttgggctaactgcgattaatttcagtaactat
15 ataccttttattccagaacataaagttactgaaactggcgactttggaggttggcaagggt
atttatgctggagtttcaatgatttttttagcttatattgggttttgactctattgctgct
aattcagctgaagcgattaatccacagaagacaatgcctagaggaatcttagggctcactc
atagtagcaattgtattgttggccgtagcacttgttcttgttggcatgttccactac
tctcaatacgttgataatgcagagccagtaggttgggcattacgagaaaagtggtcatggt
20 attattgctgcaattgttcaagcaatttctgtcatcggtatgttcaactgcattaatcggt
atgatgcttgcagggttcacgtctattatattcatttggacgagatgggttactcccttct
tggttaagtcaattgaatcacaacatttacctaatacgagcacttgtcatacttacaatc
attggcgtagttatcggtcaatgttcccggttgccttcttagcacaattgatttccgca
ggtacccttgttgcattcatgttgtgtcactagcaatgtatcgattaagaaaacgtgaa
25 gggaaagatttacctaagccagagtttaaattaccttttatatcctattttgcctgcaatt
acatttatattagtagtagtattttggggattaaagtttgaagctaagttgtatata
ctgatatgggtttattgtaggatataattatttatttatttgaattagacatttccaaa
aagaatgatgaagaagcgtagtcaagtacctaagagaataa

30 Sequence 368

VAGLVAFTYAEMASTMPFAGSAYSWINVLFGEFGWVAGWALLAEYFIAVAFVASGFSAN
LRGLIAPLGISLPSLSNPFSGNGGVIDIIAAVVIILTALLSRGMNEAARMENVLVILK
VLAIILFVIVGLTAINFSNYIPFIPEHKVTETGDFGGWQGIYAGVSMIFLAYIGFDSIAA
NSAEAINPQKTMPRGILGSLIVAIVLFVAVALVLVGMFHYSQYADNAEPVGWALRESGHG
35 IIAAIVQAIISVIGMETALIGMMLAGSRLLYSFGRDGLLPSWLSQLNKHLPNRLVILTI
IGVVIGSMFPFAFLAQLISAGTLVAFMFVSLAMYRLRKREGKDLKPEFKLPLYPIILPAI
TFILVLLVFWGLSFEEKLYTLIWFIVGIIYLIYGIRHSHKKNDEEAYQVPRE*

Sequence 369

40 Contig_0470_pos_7576_7145,

putative peptide of unknown function

atgcattgggttaaaaaatattttatcatttattatgcgcaaccacgattagcgtgatatta
cttattataactatattaatggatgcgttactacaaaacacacacttaactcagttatta
ctcaatatattgattttttaatcaaccctgatgaagtgccacaattattgaagtactgatt
45 catttaagtattggaatattgatttatctcgcctttttaattatctatcattattcaaaa
tccttgtatcatctagcatacttaccttttagtattgatatttactttgatgtatccactt
ctcggttttcttgcgcaacgtccatttttttccctttagttggaacgaatttgcaggtgg
ttagttgcacatcttttttcatcatttttaattggcgacttgtctacctatcatttcgaaa
aaaattttatga

50

Sequence 370

MHWLKIFYHLLCATTISVILLIITILMDALLQNTHTLQLLNIDFLINPDEVPTIIEVLI
HLSIGILIYLAFLIYHYSKSLYHLAYLPLVLIFTLMYPLLVLVFLAQRPFPSFSWNEFAWW
LVAHLFFIILMATCLPIISKIL*

55

Sequence 371

Contig_0470_pos_4275_3154,

is similar to (with p-value 5.0e-41)

>gp:gp|D85752|D85752_9 Enterococcus faecalis plasmid pPD1 ba
cA, bacB, bacC, bacD, bacE, bacF, bacG, bacH and bacI genes,
complete cds. NID: g2879906.

5 atgattggaataattattggtattgcatcagtaattactatcatgtcgttggggaacggt
tttaagaagtcaacgactgagcaattcaatgatgctggtgctggtaaaaatcaagcttca
atttcttacatgacagaaaatatggaagcgctaaaaataatccatttaagcaagaggat
atgagtgttgttgacaggttaattggtgttaagagtgcataaagtaaaagaggataaggat
10 agcacatattcagtcaaaattacgaatacacatggcagtagtgatgctagtttaaaaaag
gttgataaactgacagatgtagatgaaggaaaaggatttacgaatgatgataatgaagt
ctagaaaaagtagccgttatagataaaaaaattgctaaaaaagtattcaataatcaggca
atgggtcaatctatttatataaatggagaagggtttaaagtcgtaggcgtctctgaaagc
tcagaagtcgatgaaagtgggatgcctattgagtcattaattcaataaccttcaaaaaca
tttaataaatatatgggcaatttgacacaaggtatgcctcaattattagttacagttgaa
15 aaaggttcagataagaagacgtaggtaaaaaggtcgaaaaagtgttgataaaaaagga
actggcgtatctgaaggtcaatatagttatgaagataatgaagcggatgataaaacgata
ggttcagtccttagacacgattacttactttgtcgcagctgttgccgggaatatcactcttt
attgcaggtattggtgtgatgaatgtcatgtatatttcagtcactgaacgaacagaagag
attgcaatttcgctgcgtcatttggcgctaaaagggtcgagatattgaaatacaattcttagta
20 gaaagtgttgtgttatgtctcataggtggtatcatcggaattaattctaggtattattatt
gctacattgattgatctcgtgacacctgaaatgggtaagagttccgtcagtcaggttcc
gtcatcctagctgtaggtgtatcaacattgataggcatcattttcggttgataacctgca
cgttcagcttctaaaaaagaattaattgatattattaaataa

25 Sequence 372

MIGIIIGIASVITIMSLGNGFKKSTTEQFNDAGAGKNQASISYMTENMEAPKNNPFKQED
MSVVEQVNGVKSARKVEDKDSTYSVKITNTHGSSDASLKKVDKLTVDDEGKGFNDNEV
LEKVAVIDKKIAKKVFNNQAMQSIYINGEGFKVVGVSSESSEVDESGMPIESLIQIPSKT
FNKYMGNLTQGMPLVTVKEGSDKKDVGKKVEKVLNKKGTGVSEGQYSYEDNEAVMKT
30 GSVLDTITYFVAAGISLFIAGIGVMNVMIYSVTERTEEIAIRRAFGAKGRDIEIQFLV
ESVVLCLIGGIIGLILGIIATLIDLVTPEMVKSSVSLGSVILAVGVSTLIGIIFGWIPA
RSASKKELIDIK*

Sequence 373

35 Contig_0471_pos_563_1228,
is similar to (with p-value 3.0e-83)
>gp:gp|AF068904|AF068904_2 Staphylococcus aureus cell divisi
on protein FtsZ (ftsZ) gene, partial cds; YlmD (ylmD), YlmE
(ylmE), YlmF (ylmF), YlmG (ylmG), and YlmH (ylmH) genes, com
40 plete cds; and cell division protein DivIVA (divIVA) gene, p
artial cds. NID: g4009490.
atggcgagatatatcagtgacagtgacacatcatattacacatcatcaagatatcttagcg
aatcttattggttatccaagagatgaatgggtttttcctatacaaacacatgatagtcgt
atcgttgaagttacaagtgaacataaaggaaacaaatattgatgaactaactgatgattta
45 catggcatagatggaatgtatacttttgattctcacattcttcttactatgtgttatgcg
gattgcgtacctgtatatatttttatagtgaaaccacatggatatataggattagcacatgca
ggttggcgaggaacatatgggtcaaatagtaaaagaaatgctaaaaaaagtggattttgat
tatgaagacttaaaagattgtaattggtccagcaacttcaaattcttatgaaatcaatgat
gatataaaaaataaagtttgaggaattaaccattgattcaactttatatattgagaccaga
50 ggtaaaaatcaacatgggtattgatttgaaggaaggctaacgcacttcttctagaagaagct
ggagttccatcaaaaaacatatagcttacggaatatgcaacttcagaaaacttagattta
ttcttttcatatcgtgttgaaaaaggacagacgggacgtatgttagcatttattggacgg
aagtaa

55

Sequence 374

MARYISDSAHHITHHQDILANLIGYPRDEWVFPIQTHDSRIVEVTSEHKGTNIDELTDDL
HGIDGMYTFDSHILLTMCYADCVPVYFYSEPHGYIGLAHAGWRGTYGQIVKEMLLKVD

YEDLKIVIGPATSNSEYINDDIKNKFEELTIDSTLYIETRGKNQHGIDLKKANALLLEEA
GVPSKNIYVTEYATSENLDLFFSYRVEKGQTGRMLAFIGRK*

Sequence 375

5 Contig_0471_pos_1254_1922,
is similar to (with p-value 2.0e-92)
>gp:gp|AF068904|AF068904_3 Staphylococcus aureus cell divisi
on protein FtsZ (ftsZ) gene, partial cds; YlmD (ylmD), YlmE
(ylmE), YlmF (ylmF), YlmG (ylmG), and YlmH (ylmH) genes, com
10 plete cds; and cell division protein DivIVA (divIVA) gene, p
artial cds. NID: g4009490.
atggatgttaaagagaatcttgctaagattgaaaaggaaattgatgcaagcattaaaaaa
agtgcgcatttcagcaaacctcacgtgattgcagtaacaaaatatgttacaatagagcga
gctagagaagcgtataaaagtagggataagacatttcggtgaaaatcgattagatggattc
15 aaagagaagaaagaatctctaccaagcgatgttaaattacatttcattggttctttacaa
tcaaggaaagtaaaagatattataaatgaagtcgattattttcatgcttttagatcgttta
agtctagctaaggagattaataaaaagagcaaatcatgttataaaatgtttcttacaagta
aatgtttctggagaagaatctaaacatggcatagctcttgaagaagtgaatcaatttata
aatcaaatataagaatagaaaatatccaaattattggattaatgacgatggcaccattg
20 actgatgatttatcgtagacataagaaatttatttaaagaattaagacataaaaagaatgaa
attcaacaattcaatttagcacatgcaccttgtagagaattatctatgggaatgagtaat
gattatcaaattggcagttgaagaaggtgcaacctttgtcagaattgggactaaacttgta
ggagaatag

25 Sequence 376

MDVKENLAKIEKEIDASIKKSAHSAQPHVIAVTKYVTIERAREAYKVGIRHFGENRLDGF
KEKKESLPSDVKLHFISLQSRKVKDIINEVDYFHALDRLSLAKEINKRANHVIKCFLOV
NVSGEESKHGIALEEVNQFINQIKEYENIQIIGLMTMAPLTDDL SYIRNLFKELRHKRNE
IQQFNLAHAPCTELSMGMSNDYQMAVEEGATFVRIGTKLVGE*

30

Sequence 377

Contig_0471_pos_2176_2529,
is similar to (with p-value 5.0e-45)
>gp:gp|AF068904|AF068904_4 Staphylococcus aureus cell divisi
35 on protein FtsZ (ftsZ) gene, partial cds; YlmD (ylmD), YlmE
(ylmE), YlmF (ylmF), YlmG (ylmG), and YlmH (ylmH) genes, com
plete cds; and cell division protein DivIVA (divIVA) gene, p
artial cds. NID: g4009490.
atgaataataattcaaaaaataattctagaacgttgtaacaatgaaccaagcatcacaa
40 tcatatgccgctcaggaaagttcaaaaatgtgtctgttgtaaccacgtgtcttttcagat
actcaagatattgccgacgaattaaaaaacagacgtgcaacttttagtaaatttacaacgc
attgatcaagtatcagcaaagcgtattattgattttttaagtggtagcggtatacgcaatt
ggtaggatattcaacgcgtgggtactgatattttcttatgcacacctgataatgttgaa
gtagccggtagtagtaactgatcacatcgagaatatggagcaacactacgaataa
45

Sequence 378

MNNNSKNNSRVVTMNOASQSYAAQESSKMCLFEPRVFSDTQDIADELKNRRATLVNLQR
IDQVSAKRIIDFLSGTVYAIGGDIQRVGTDIFLCTPDNVEVAGSITDHENMEQHYE*

50 Sequence 379

Contig_0471_pos_3052_4308,
is similar to (with p-value 1.0e-39)
>gp:gp|AF015775|AF015775_17 Bacillus subtilis YodA (yodA), Y
odB (yodB), YodC (yodC), YodD (yodD), ABC-transporter (yodE)
55 , permease (yodF), proteinase (ctpA), YodH (yodH), YodI (yod
I), carboxypeptidase (yodJ), purine nucleoside phosphorylase
(deoD), YodL (yodL), YodM (yodM), YodN (yodN), YodO (yodO),
YodP (yodP), acetylornithine deacetylase (argE), butyrate-ac
etoacetate CoA transferase (yodR), butyrate acetoacetate-CoA

transferase (yodS), YodT (yodT), CgeE (cgeE), CgeD (cgeD), CgeC (cgeC), CgeA (cgeA), CgeB (cgeB), YzxA (yzxA), UDP-glucose epimerase (yodU), YodV (yodV), and YodW (yodW) genes, complete cds; and YodZ (yodZ) gene, partial cds. NID: g2415383

5 . >gp:gp|Z99114|BSUB0011_133 Bacillus subtilis complete genome (section 11 of 21): from 2000171 to 2207900. NID: g2634230.

atgggatatcaattggataaacgtcaatttgaatttttagatatgctagttagatttaata
actgaaagtcacacgtgacgtaatacagatccattgcaagatgaaatcgaaacgttactt
10 aaacaactggatttttcaatacagagagaacagttatacgacaatgatagtgtgatagta
gctaccttaaaagggcacaaatcctaaagcgccaaaactgatattgaatggacatgttgat
gtagcttctgtagatgacgatcaatattggcagtatccaccttttaacttaccaaaaa
gaagaatgggttatacgggtcggttagcgatatgaaaggtgggtatgtcttcattattc
tacgtcttgagcaattacatcaagaggggcaacgtccagaaggtgatattattgttcaa
15 tcagtagtcggtgaagaagtaggtgaagcaggaactaaacgtgcatgtgaaataggacct
aaaggtgacttagcccttgtcttagatagcagtgagaatcaagcacttgggcaaggtggc
gtgattaccggatggattacagttaaaagtaaaatacaatacatgatggtgcgcgtagt
caaacgatacatgctggtgggggctgtttggtgcaagtgccattgaaaaatgacaaag
gtgattccatcgcttaatagaacttgaaggcattgggctgtcatgaagaagagccctgga
20 atgcctccaggtgcgaataacaattaacccagctgtcatagaaggtggacgtcaccctgca
tttattgcagatgaatgtcgattatggattactgttcattacttaccgaacgaaagtta
gaatctgtagttaatagaatagagcaatatttaataaaggttcagaagcagatgtatgg
ctcagagagaatccacttgaatttgaatgggggtggtacatccatgattgaggataaagga
gaaatcttcccaagtttcaactgttccgacacatcatccaggttttaagcaattagaagaa
25 gcacatgaacatattcataataaaaaagcttgaacatggtatgagtacaactgtaactgat
ggaggttggacagcacattttggcattcccacgatatttatggcccaggtagtttagaa
gaggcacatagtgtagatgagaaaaataaaagcaaaaggaatttagctcaatatagtgtgt
ttatatacattttttaaaagagtggtatgcacaccacaaatcctataaatcatcatag

30 Sequence 380
MGYQLDKRQFEILDMLVRFNTESPPGRNTDPLQDEIETLLKQLDFSIOREQLYDNDSVIV
ATLKGHNPKAPKLIILNGHVDVASVDDQYWOYPPFKLTNKEEWLYGRGVSDMKGGMSSLF
VYLEQLHQEGORPEGDIIVQSVVGEVGEAGTKRACEIGPKGDLALVLDTSNQALGQGG
VITGWITVKSNTIHDGARSQTIHAGGGLFGASAIEKMTKVIQSLNELERHWAVMKKSPG
35 MPPGANTINPAVIEGGRHPAFIADECRLWITVHYLPNESYESVNEIEQYLNKVAEADVW
LRENPLEFEWGGTSMIEDKGEI FPSFTVPTHHPGFKQLEEAHEHIHKKLEHGMSTTVTD
GGWTAHFGIPTILYGPSLEEAHSVDEKIKAKELAQYSDVLYTFLKEWYAHQPQSYKSS*

Sequence 381

40 Contig_0471_pos_8552_0,
is similar to (with p-value 4.0e-39)
>sp:sp|P39640|YWFD_BACSU HYPOTHETICAL OXIDOREDUCTASE IN ROCC
-PTA INTERGENIC REGION (EC 1.-.-.-). >pir:pir|S39737|S39737
hypothetical protein - Bacillus subtilis >gp:gp|X73124|BSGEN
45 R_83 B.subtilis genomic region (325 to 333). NID: g413923. >
gp:gp|Z99123|BSUB0020_68 Bacillus subtilis complete genome (section 20 of 21): from 3798401 to 4010550. NID: g2636240.

atggaacgttttagaaaacaaaatcgacgtgattactggtgcgagtactggtattggacaa
gcacgcggccgtggcggttagcaaaagaaggagcacatgtgttagcgcttgatatatcagat
50 caattagaagaaactgtgcagtcattataatgataatggtgggaaagcaactgcatacgc
gtagacatttcagatgataaacaagtcaaacattctcagaaaaatagcacaagaattt
ggacatgtagatgttatttttaacaatgcggtgtagataatggcgccggacgtattcat
gaatatccagttgaagtgtttgataaaattatggctgttgatatgagaggaactttttta
tgaactaaatttttattaccttttaattgatgaacaagggtggttctattattaatacagct
65 tcatctctctgggcaagctgcggtttataccgttcagggtataatgctgctaaggcggt
gtcattaattttacaaaatctatcgctatagaatatggacgtgaaaaatattcggtgcta
gctatagcacctggaacaatcgaaacaccacttgttgataatttagcaggtacatcagat
gaagaagccggacaaacattccgagaaaaatcaaaaatgggtaacaccattaggtcgacta
ggaacaccggatgaagtgtggaaactttagtccttttttagcttccgatgatagttcattt

ataactggtgaaactattcgtattgatgggtggcgtgatggccttatacatggctacaacac
gcattttcattttttgtcgttggttttttctattttcttgtagcgtatattggttaccga
tatgttctacttttttattcattgcttgctcacctccttaagcatttcactcttcattaat
acgttcttctttaatc

5

Sequence 382

MERLENKIAVITGASTGIGQASAVALAKEGAHVLAALDISDQLEETVQSINDNGGKATAYR
VDISDDKQVKQFSEKIAQEFQGHVDVIFNNAGVDNGAGRIHEY PVEVFDKIMAVDMRGTF
VTKFLLPLMMKQGGSIINTASFSGQAADLYRSGYNAAKGGVINFTKSIAIEYGRNIRAN
10 AIAPGTIETPLVDNLGTSDEEAGQTFRENQKWVTPGLRLGTPDEVGKLVAFSLASDDSSF
ITGETIRIDGGVMAYTWLQHAFSFFVVVFSYFLVTYIGYRYVLLFYSLLVTSLSISLFIN
TFFFNX

Sequence 383

15 Contig_0472_pos_3883_3548,

putative peptide of unknown function

atgggtgctcgtacaatttccaccttggttgattgtaacgtccaaaatataaattacatc
ttatatgtgagaaaacaattaactgatattccgatgagcattgaatttagacatcaatca
tggtttgacaatcagtataaagaacaaactttatccttcttaacacacatcaaatcatt
20 catgcatgtgtagaactcaagttaaagaggggagcgttccttttagtaaataggatt
actagtgaattgcttttgtagcttatcatggacgtaatcattatgggttgactaaaaaa
gatatgactgatcaagaatggcgagatgtaagataa

Sequence 384

25 MVLVQFPWFDCNVQININYLYVRKQLTDIPMSIEFRHQSWFDNQYKEQTLSTLQHQII
HAVVDEPQVKEGSPVLNRTSEIAFVRYHGRNHYGWTKKDMTDQEWDRDVR*

Sequence 385

30 Contig_0473_pos_900_2051,

is similar to (with p-value 8.0e-75)

>sp:sp|Q01444|YFF1_MYCMY HYPOTHETICAL PROTEIN IN FFH 5'REGIO
N (FRAGMENT). >pir:pir|S35480|S35480 hypothetical protein 1
- Mycoplasma mycoides (SGC3) >gp:gp|M91593|MYCSRPM54A_1 Myco
plasma mycoides SRPM54 gene, complete cds. NID: g150208.

35 atgaagcatgaacaagaattagaacgcatctccggtctcactcaagaagaagctgtgaaa
gaacagcttcaaagagttgaagaagaactgtcacaagatattgcaatacttgtaaagaa
aaagaaaaagaagcgaaagaaaaagttgataagacagctaaagaattacttgctacaact
gtacaagattagcagccgaacatacaactgaatcaactgtttcagtcgtaaatctgcct
aacgatgaaatgaaaggtcgtatcataggtagagaaggtagaatatacgacacattagaa
40 acactactggcatagatttaattattgtgacacaccagaagcagttattttatcaggt
tttgacccaatttagacgtgaaattgctagaactgcactagtttaatttggtttctgatgga
cgtattcatcctggacgtattgaagatatggtcgaaaaagctagaaaggaagtagacgat
atcatttagagatgctggagaacaagctacctttgaaataaatgtacacaatatgcatcct
gatttagtgaaaattttgggtcgattaaattatcgaaactagttatggtcagaatgtactt
45 aaacattcaattgaagttgccacctttcaggtatgcttcgacgagaattaggagaggat
gttacttttagctaaacgtgctggattattacatgatgttggtaaagccattgatcatgaa
gttgaaaggtagtcacgtagaaataggtgttgatttagctaaagaaatataatgaaaataac
ataattattaatgctattcactcacatcatgggtgatgttgaaaccaacctctatcatttct
attttagttgcagcagctgatgcattatcagcagcgcgaccaggtgcacgtaaagagaca
50 cttgaaaatttatattagaagacttgagagactcgaaacgttatctgaaagttatgatggg
gtagaaaaagcatttgcatacaagctggtagagagattcgtgtagtcgtctcacctgaa
gaaattgatgatttaaaatcatatagattggcaagagatattaagaaccaaattgaagaa
gagttacaatatcctggacatatcaaagtgcagttgttcgagagactagagcaatagaa
tatgctaaataa

55

Sequence 386

MKHEQELERISGLTQEEAVKEQLQRVEEELSQDIAILVKEKEKEAKEKVDKTAKELLATT
VQRLAAEHTTESTVSVNLPNDEMKGRIIGREGRNIRTLETLTGIDLIIDTPEAVILSG
FDPIRREIARTALVNLVSDGRIHPGRIEDMVEKARKEVDDIIRDAGEQATFEINVHNMHP

DLVKILGRNLNYRTSYGQNVLKHSIEVAHLSGMLAAELGEDVTLAKRAGLLHDVGKAIDHE
 VEGSHVEIGVELAKKYNENNIINAIHSHHGDEPTSIISILVAAADALSAARPGARKET
 LENYIRRLERLETLSesydgvekaFAIQAGREIRVVVSPEEIDDLKSYRLARDIKNQIEE
 ELQYPGHIKVTVVRETRAIEYAK*

5

Sequence 387

Contig_0473_pos_2519_3313,

is similar to (with p-value 4.0e-37)

>sp:sp|P47488|Y246_MYCGE HYPOTHETICAL PROTEIN MG246. >pir:pi
 10 r|B64227|B64227 hypothetical protein MG246 - Mycoplasma geni
 talium (SGC3) >gp:gp|U39703|U39703_14 Mycoplasma genitalium
 section 25 of 51 of the complete genome. NID: g3844835.
 atgagaatattgtttataggtgacatcggttggttaaagtggcaggaaaatgattactact
 tatttacctaaaattaaacaaacttatcacccaacagtttctatagtaaacgctgaaaat
 15 gccgcacacggtaaaaggattaacagaaaaaatttacaacaacttttgagagaaggcgtg
 gatttcatgactatgggtaatcatacatatgggtcaaagagaaatttacgattttattgat
 gatgctcatcgaatgggtgagacctgcaaattttctctgatgaagctccaggaaacaggatg
 agaataataaaaaattaacgatattaaattggctattattaattacaaggcgttcattt
 atgcaagacattgatgatccatttaaaaaggctgaccagctaatacgaagaagctcaaaaa
 20 tctacaccatataatattgttagattttcatgctgaaactacatctgaaaaaaatgctatg
 gggttggtatttagatggtagagttagcgctgttggttggtactcacacacataattcaaact
 tctgatgatcgatattacctcatggcacaggatataatcacagatgtcgggatgacaggt
 tattacgatggatttttaggtatcaatagagatgaagttattcaacgttttattactagt
 ttgccacaaaggcatgttggtccagatgatggcgaggcgtattatcaggaggttatcata
 25 gatttagataaagaaggtaaaaacgactcaataaaaaagactgttaataaatgaggacat
 cctttccaaatttaa

Sequence 388

MRILFIGDIVGKVRKMITTYLPKIKQTYHPTVSIVNAENAAHGKGLTEKIYKQLLREGV
 30 DFMTMGNHTYQREIYDFIDDAHRMVRPANFPDEAPGTGMRIKINDIKLAIINLQGRSF
 MQDIDDPFKKADQLIEEAQKSTPYIFVDFHAETTSEKNAMGWYLDGRVSAVVGTHTHIQ
 SDDRILPHGTGYITDVGMTGYDGLGINRDEVIQRFITSLPQRHVVPDDGRGVLSGVII
 DLDKEGKTTQIKRLLINEDHPFQI*

35 Sequence 389

Contig_0473_pos_3372_5186,

is similar to (with p-value 4.0e-83)

>pir:pir|S22396|S22396 pyruvate synthase (EC 1.2.7.1) - Halo
 bacterium halobium >gp:gp|X64521|HHFEROXI_1 H.halobium gene
 40 for pyruvate:ferredoxin oxidoreductase. NID: g43497.
 gtggtttattttgttatcatagagtatgaattaataccacaggaggcatgtgatatgaaa
 tcacaaatatcatggaaagtggcggtcagcaaggcgaaggattgaatctaccggtgaa
 atctttgctactgcatgaatagaaaagggtattttttgtatggatatagacacttttct
 agtcgtataaaaagggtggccataactaataaagataagagtttcaaaatcgctgtgcat
 45 gcgattagtgtgatttggatatactcattgtctttgaccaggaaacgattgaattaat
 catcatgaaatgagagaagatagattataaattgcggatgctaaagcaaaacccaaaag
 ccagagaactgtgtggctcaattaattgagttaccattcactagcacggcaaaaggaactt
 ggaacagcattaatgaagaatatgggtggcaattgggtgcgacatctgcactgatggattta
 aatacatcaacttttgaaactttaatcgataacatgttttcaaaaaaggtaataaagtc
 50 gttgatatgaatatacaagcccttaatatgggttatgatttaatgaagcaacaagttacc
 aacgttaatggagactttacattagagaatggtagcggtcatcctcatttatatatgata
 ggtaattagcgaatcggttaggagcaatagcagctggatcaagatttatgtccgcttat
 ccaattacgcagcttctgaaattatggaatacatgattgccaatctacctaagttgat
 ggtactgttgttcaaaactgaagatgaaatagcagcagcaacgatggcgattggagctaac
 55 tatgctggcgtacgaggctttacagcgagtgcggtccagggtctttctttaatgatggaa
 tctattggattgtctggatgactgaaacgccattagtattattaataactcaaagaggt
 ggctccttctactggcttaccacaaagcaagaacaatcagatttaatgcaaatgatttat
 ggtacccatgggtgatattccgaaaattgtcggtgctcctacagatgctgaagatgcgttt
 tatcttactatggaagcatttaatttagctgaagaataccaatgtccagtcattctgtta

agtgattttacaattatcattagggaaaacaaactgttaaaacactcgattataataaaatc
 gatatttcgctcgaggagaaataatacagtcagatatcgagagagctgaagatgataaagca
 tactttaaaagatatgcattaacagctagtggcgtatcaccacgaccaataaccaggtgtt
 aaaggtggtatacatcatgtaacaggtgttgacataatgaagaagggaagccaagtga
 5 ggcctatgaatcgtcagaatcagatggaaaaacgaatgcgcaaaactgaaagcttggtt
 atcaataatcctgtgttactcaatgaacatgaagacgaagcagatatactgtatatagga
 ttatatctactaaaggtgctattggagaaggtgcagaaagactagaacgacatggtgta
 aaagtgaatacagatgcatttcgacaattacatcctttccctaagatattgttcaacaa
 gctattaataaagcttcgaaagtaatagttgcagaacataattatcaaggacaattatca
 10 agtatttttaaaaatgaacacacaagttaatgataaattagttaatcaaacaaaatcagat
 gggaaacctttcttacccttatgaaattgaagaaaaaggtttggaattgctaaagagtta
 aaggagttggtgtaa

Sequence 390
 15 VVYFVIIIEYELIPQEACDMKSQISWKVGGQQGEGIESTGEIFATAMNRKGYFLYGYRHFS
 SRIKGGHTNNKIRVSKSPVHAISDDLDILIAFDQETIELNHHEMREDSIIIADAKAKPQK
 PENCVAQLIELPFTSTAKELGTALMKNMVAIGATSALMDLNTSTFETLIDNMFSSKGNKV
 VDMNIQALNMGYDLMKQQVTNVNGDFTLENGSGHPHLYMIGNDAIGLGIAAGSRFMSAY
 PITPASEIMEYMIANLPKVDGTVVQTEDEIAAATMAIGANYAGVRGFTASAGPGLSLMME
 20 SIGLSGMTETPLVIINTQRGGPSTGLPTKQEQSDLMQMIYGTHGDIPKIVVAPTDADF
 YLTMEAFNLAEYQCPVILLSLQSLGKQTVKTLTDYNKIDIRRGEIIQSDIERAEDDKA
 YFKRYALTASGVSPPPIPGVKGGIHHVTGVEHNEEGKPSEAPMNRQNMKMRKTESLV
 INNPVLLNEHEDEADILYIGFISTKGAIGEGAERLERHGVKVNMTMHIRQLHPFPKDIVQQ
 AINKASKVIVAHEHNYQGLSSILKMNTQVNDKLVNQTKYDGKFFLPYEIEEKGLEIAKEL
 25 KELV*

Sequence 391
 Contig_0473_pos_6639_7256,
 is similar to (with p-value 2.0e-18)
 30 >pir:pir|S41182|S41182 hypothetical protein 37.1 - phage SPP
 1 >pir:pir|S43808|S43808 hypothetical protein 38 - phage SPP
 1 >gp:gp|X67865|BSSPP1_10 B.subtilis phage SPP1 DNA sequence
 coding for products required for replication initiation. NI
 D: g472886.
 35 atggacaaattttaaatctatgacagaattaaaagaattgactaaagaaggaaaagattgg
 gaaatagagtgtgaaaatcgttctagcatagtcactatattagcattacatggcgggtgga
 attgaacctgccacaactgaattagcctatacaattgcacattgtggcgactataactat
 ttttccttttaaaggatgagaagtaaggggaataatgagttacatgtgacttccacacat
 tatgatgaccaaattgcattagatttagtgagaggtagccaaagaactgtagccatccat
 40 ggttgtgaaggtaatgaaagtgtggcttatataggaggtagtgatgacagactaattgag
 ttaatcaccgaatctcttgaagatataggaattagcgtgcgagaagcaccacatcatatt
 tctggaactcaagaaaataatattgttaatatgactcaaaccaaggaggagtgaatta
 gaactgacagctcagttaagaaaggagctattttaaaaatagaaaaagttcacgcaaaaac
 cgtgaaaataaagataattgggatgatttaattgtacgactttgtgtgatgcaatgaaaaaa
 45 gctatagaacgtgcataa

Sequence 392
 MDKFKSMTELKELTKEGKDWEIECENRSSIVTILALHGGGIEPATTELAYTIAHCGDYN
 FSFKGMRSKGNLHVTSTHYDDQIALDLVRGSQRTVAIHGCEGNESVAYIGGSDDRLIE
 50 LITESLEDIGISVREAPHHISGTQENNIVNMTQTQGGVQLELTAQLRKELFKNRKSSRKN
 RENKDNWDDLMYDFADAMKKAIERA*

Sequence 393
 Contig_0473_pos_8529_9131,
 55 putative peptide of unknown function
 atgaaaaatgtttctaaagctttgatttggtttgtataagcttcatcatctttcacgca
 atattatttggatgtggggagaaacatcaagaatactggattttatatactggcattatg
 ttaatagctggaataagttatgttttttaccaaagagacattgcattctaaacgattatta
 acttccataggcatgggtataataacgagtgctgcacttattattatacaattaattttt

tcacttatttcatcagaattatcatacgcacgtctttaatcaaagaattatcacgaacgggt
gtctacttttaaatggcaaatgctcgttactttattatttgtgataccttgtcatgaatta
tatatgagaactgttttacaaaaggaattataaaatataacttaccgaaatgggctagc
atthtaattgttgcaatatgttcaagttcattttatatacttagataattgggtgatt
5 gtattctttattttttagctcaattcattctatctcttagctatgaatatacgagacgt
attgctacgactacaattgggtcaaattgtggctatcattttattattgatattccacgga
taa

Sequence 394

10 MKNVSKALIWFVISFIIHAILFVMWGEHQEYWYLYTGIMLIAGISYVFYQRDIASKRLL
TSIGMGIITSVALIIQLIFSLISSELSYASLIKELSRTGVYFKWQMLVTLFVIPCHEL
YMRVLQKELIKYNLPKWASILIVAICSSSLFIYLDNWWIVFFIFVAQFILSLSYEYTRR
IATTTIGQIVAIILLIFHG*

15 Sequence 395

Contig_0474_pos_2713_1850,
putative peptide of unknown function
gtgttaataatgaatgtcttccaaatgagagataaattgaaagcgcggtttaaaacattta
gacgtagaattcaagtttgatagagaagaagaaacgttacgtattgtaagaattgacaat
20 cacaaaggtgtaacgattaaacttaacgctatcgctcgaaaatatgaagaacaaaagaa
aaaattatagatgaaatttggtattatgtcgaggaagcaatcgctcagatgggtgatgaa
gtgattaataatgttgaggacatacaaaattatgccggttataagagctacaagtttcgac
aaagaaactaaggaaggtcatgcatttgtgttaacagacatactgctgaaactaatata
tattacgctcttgatctagggaaatcttatcggtcaatagatgaaaatatgttacaacg
25 ttaaaatttaactgctcaacaagtgaagaaatgtcactatttaattgttcgtaagttagag
tgtcgctatagtacggatgaagttaaaggtaatattttttacttcatcaacacaaatgat
ggatatgatgcaagtcgtatttttaataacttcttttttaaatcatattcaacaccaatgt
gaaggtgaaatgcttgttggtgtgccacatcaagatgtatttaattcttgcagatattaga
aataaaacaggttatgatgttatggctcatttgactatggaattctttactaaaggactt
30 gttccgattacttctttatcatttgggttatgataacggacatctagagccaatatttatt
ttggggaaaaataataaacaacaaaagagatcctaacgttattcaacgtttagaagcgaac
agaaaaaaattcaaaaaagattaa

Sequence 396

35 VLIMNVFQMRDKLKLKHLDFEKFDFREEETLRIVRIDNHKGVTIKLNAIVAKYEEQKE
KIIDEICYVEEAIAQMGDEVINNVEDIQIMPVIRATSFDKETKEGHAFVLTEHTAETNI
YYALDLGKSYRLIDENMLQTLNLTAQQVKEMSLFNVRKLECRYSTDEVKGNIFYFINTND
GYDASRILNTSFLNHIHQCEGEMLVGVPHQDVLILADIRNKTGYDVMALHTMEFFTKGL
VPITSLSFGYDNGHLEPIFILGKNNKQKRPNVIRLEANRKKFKKD*

40

Sequence 397

Contig_0474_pos_0_1210,
45 putative peptide of unknown function
atgagctggtttgataaattatttggcgatgacaacggttcgaatgacgatttggtacgc
aaaaataaaaaatagacgtcagtcctcagcaatcaaaacaaaataatcaagactcattactg
cctcaaaataatgatatttatagtcgaccaagaggtaaatttagatttccaatacaagtt
tctgaaaatgaatatacgcaaaaaaatgaaaattataatgaacataaccaagaagaaaca
50 aacgatataatgagatcatataaccagcatgataatcctgaatttgattcttctggtaaa
agacatcgacgcccagcgccagcgtattcaaaacacgatcaatctaagattacacaacaa
aagcaatttgcagataaacaattatacaaaataataacagtggttttaatacaaaacgacaat
aagaaatcttcacaacaacgtaaatcaatacaatctgaaaatatcaaaacaaagcaaac
actaagaatacgtcgacatctcctgaatttacatatttaaatcatagtttttaaatcaagc
55 gaggtaccctcagcgatttttgggtacaaaaaacgaagaccgattgagaatgggtgtcata
ccgccagaacataaggaattaaatgataaagagattgttcaacaggatgaagtctcgcat
tcaacgaaatcaatagatgcatcaaaaaatgtttctaataagtaacgataacaatattgaa
aaaaatcaacagaaaaaacaacaaactgctcaaaactgagtcacatcagaaaatag
cataatgttgaaaagtcgaattatcaaaactactaagcgtaaaacaccaaattactctaaa

gtagataatacgattaatattgaaaatatctatgcttcacaaattgtagaagaaatcaga
 agagaagagaacgtaaaagtctacagaaacgcgcttttaagaaagccttacaacaaaaa
 cgtcaacaaaatcaacagtcagaagaggattcaattcaaaaagctattgatgaaatgtat
 gctaagcaagcccaacattacacaggcgaaagttcattggatttagaaaatgaaagtaat
 5 caagattcgatcatctaatagtctagagaaacaatcaaatagcagcaacattgacaataa
 gaagcccaaaaataacacacctttattttaactacgaagaattgacttagatacgacatca
 gatgtTCTTC

Sequence 398

10 MSWFDKLFGGDDNGSNODLLRKNKNRRQSQQSKQNNQDSLPPQNNDIYSRPRGKFRFPQV
 SENEYTKKNENYNEHNQEETNDIMRSYNQHDNPEFDSSGKRHRRRRQAYSKHQSKITQQ
 KQFADNNYTNNSNVFNQNDNKKSSQQRKSIQSENINKNKANTKNTSTSPEFTYLNHSFKSS
 EVPSAIFGTTKRRPIENGVIPEHKEKLNDEIVQQDEVSHSTKSIDASKNVNSNDNNIE
 KNQOKKQQTTAQTESSENMHNVEKSNYQTTKRKTPNYSKVDNTININIENIYASQIVEEIR
 15 RERERKVLQKRRFKKALQKQKQNNQSEEDSIQKAIDEMYAKQAQHYTGESSLDLENESN
 QDSSSNSLEKQSNSSNIDNKEAQNNTPLFNYEEDLDLTTSDVLX

Sequence 399

Contig_0475_pos_6773_7180,
 20 is similar to (with p-value 3.0e-46)
 >gp:gp|Y13384|LLNISZ_1 Lactococcus lactis nis2 gene and 3 O
 RF's. NID: g3157416.
 atgatgaaaaataaattacattaaaagagaatctattttatcggtcattgctgtttggt
 cttttttttggtgctggaatctcatttttccaattcacttaggtcaaactgcgggggca
 25 aatgtatggacgcgaatttaggatttcttatcagcgctatcgactaccttttttagga
 attatagcgataggtgtatctaaaacaaacggggtctttgaaatttcctcaaggataagt
 aaaatatatggttatttgttcacaattggcttgtatcttgttataggtccgttttttgcg
 ttgccaaagacttgcgacgacgtcatttgaatagcattttcaccatttatttcattctggt
 acggcccaagcgttgttgcatttttttagtattttattcttcggagtag
 30

Sequence 400

MMKNKLTLENLFIGSMLFGLFFGAGNLIFPIHLGQTAGANVWTANLGLITAIGLPFLG
 IIAIGVSKTNGVFEISSRISKIYGYLFTIGLYLVIGPFFALPRLATTSFEIAFSPFISSG
 35 TAQALLLFLVFYSSE*

Sequence 401

Contig_0475_pos_7273_8133,
 is similar to (with p-value 1.0e-49)
 >sp:sp|P54104|BRNQ_LACDL BRANCHED CHAIN AMINO ACID TRANSPORT
 40 SYSTEM CARRIER PROTEIN. >pir:pir|S60180|S60180 branched-cha
 in amino acid carrier brnQ - Lactobacillus delbrueckii >gp:g
 p|Z48676|LDBRNQGN_1 L.delbrueckii brnQ gene for branched-cha
 in amino acid carrier. NID: g732812.
 gtgcttgcatctatccgtcctatgggtggaattagtcgcatgcgcagtaagtgtgattat
 45 agcaatagcgtgttactcaaagggtttatcgatggatataatacattagacgctttggca
 tcattagcatttggattatcattgttactacaattaaaaagttggggattactaatccg
 aatacaatcgctaaagaaacttttaaaatcaggtacgattagtagtattatagctatgggcgtt
 atttatactttattagcttttaattgggtacgatgagtttaggtcgttttaaagtaagtga
 aatgggtggtatttgcgttgcagattgcacaacattatttaggggattacggaattatt
 50 attttgcactaatcatcattgtggcatgtctgaaaacagcaataggattgatcacagcc
 ttttcggaacattttacagagttattccctaaatctaactatctttggttagctactggg
 gtgagtatatttagcttgtatatttgcataatgtagggttaacaaaaattattatgtattca
 acaccagtgttgatgttcatttatccttttagcgattactttaattttattagcattactt
 agtccattatttaaacatttctaaaattgtctatcgatttacaacattatttacaatgggtg
 55 gcggcatttgtagatgggtgtgaaagcaagtcagagttctttgttaatacaaaaatttgca
 caaacaatcattggatttgggtgaaaattatctccattctttaacattggatgggatgg
 attgttccagcacttattggtttcattattgggtattattgtatactttatgactgctaaa
 aatcgtccacgtacaataa

Sequence 402

VLA FIRPMGGISHAPVSADYSNSVLLKGFIDGYNTLDALASLAFGIIIVTTIKKLGITNP
 NTIAKETLKSGTISIIAMGVIYTLALMGTM SLGRFKVSENGGIALAQIAQHYLG DYGII
 ILSLIIIVACLKTAIGLITAFSETFTELFPKSNYLWLATGVSILACIFANVGLTKIIMYS
 5 TPVLMFIYPLAITLILLALLSPLFKHSKIVYRFTTLETFMVAAFVDGVKASPEFFVNTKFA
 QTIIGFGENYLPFFNIGMWIVPALIGFIIGIIVYFMTAKKSSHVQ*

Sequence 403

Contig_0475_pos_5847_5449,
 10 putative peptide of unknown function
 atgccacacctaaggtacaaatgctgttgatatttttagttgattttgtaaatgaaatgaaa
 caagaatataaaaaatttaaagaacatgataaagtagacaggttagacgctgttccaatg
 attgagaacacatctccacagaaaaattggtgaagaagaatcacatatctactctggattt
 gtaatgttaaactctgtattcaatggtggtgaaacaagttaattctgttcctcataaagcg
 15 acagctaaatataatgaagaactgttccagaatatgacagtactttcgtgaaggatttta
 tttgaaaaagtcattcgtcatgtggcggaagattatttaactgtagatatacctagcagt
 cacgatccagtggaagtgtatcgttgagatttaattaa

Sequence 404

20 MPH LGTNAVDILVDFVNEMKQ EYKNIKEHDKVHELDAVPMIEKHLHRKIGEEESH IYSGF
 VMLNSVFNGGKQVNSVPHKATAKYNVRTVPEYDSTFVKDLFEKVIRHVGEDYLTVDIPSS
 HDPVASDRWRFN*

Sequence 405

25 Contig_0475_pos_4351_2876,
 is similar to (with p-value 4.0e-90)
 >gp:gp|AF006665|AF006665_31 Bacillus subtilis 168 region at
 182 min containing the cge gene cluster. NID: g2529445. >gp:
 gp|AF015775|AF015775_7 Bacillus subtilis YodA (yodA), YodB (
 30 yodB), YodC (yodC), YodD (yodD), ABC-transporter (yodE), per
 mease (yodF), proteinase (ctpA), YodH (yodH), YodI (yodI), c
 arboxypeptidase (yodJ), purine nucleoside phosphorylase (deo
 D), YodL (yodL), YodM (yodM), YodN (yodN), YodO (yodO), YodP
 (yodP), acetylornithine deacetylase (argE), butirate-acetoac
 35 etate CoA transferase (yodR), butyrate acetoacetate-CoA tran
 sferase (yodS), YodT (yodT), CgeE (cgeE), CgeD (cgeD), CgeC
 (cgeC), CgeA (cgeA), CgeB (cgeB), YzxA (yzxA), UDP-glucose e
 pimerase (yodU), YodV (yodV), and YodW (yodW) genes, complet
 e cds; and YodZ (yodZ) gene, partial cds. NID: g2415383. >gp
 40 :gp|Z99114|BSUB0011_121 Bacillus subtilis complete genome (s
 ection 11 of 21): from 2000171 to 2207900. NID: g2634230.
 atgaatgatcatcaaaaaaatcatgcaacatctcaagatgataacacaatgtcaacacca
 tctaagaatagcaagcatataaaaaattaaattatggcatttcataactcgttattttgggt
 attattcttttaacatccatcattactgtagtagtcaacaatttttaattagccatcaaaaa
 45 agtgggtttaataaagaacaacgtgcaaatttaaaaaaaattgaatatgtctatcaaaaca
 cttaataaaagattattacaaaaagcaaagttctgataaattaaactcaatctgccatagat
 ggtatgggttaaaagaacttaaaagatccatattcagaatatatgactgctgaagaacaaaa
 caatttaataagaggtgtatcaggtgatttcgttggcataggtgctgaaatgcaaaaagaaa
 aatgaacagataaagtggttactagcccaatgaaggattcaccagcagaaaaagctggtatt
 50 caacctaaagatatcgtcacacaagtgaatcatcattcggtagtcggtaaacacttgat
 caagttgttaaaatggtccgcggaaaaaaggaacatatgttactttaactataaaacgt
 ggttcgcaagaaaaggatattaagattaaacgcgataaccattcacgttaagagtgtagag
 tatgagaagaaaaggcaatgtaggcgtactaacaatcaataaattccaaaagcaatacttct
 ggtgaactcaaatctgcaatcatcaaaagcgcataagcaaggcatccgtcatatcatttta
 55 gatttgagaaataatccgggggggttattagatgaggcagtcagatggctaacaatcttt
 attgataagggaaataactgtcgttcaattagaaaaaggaaggataaggagaattaaaa
 acttctaatacaagcactaaaacaagcaaaagatatgaaagtatccatcttagttaatgag
 ggatcagctagtgcttcagaagtgtttacaggtgctatgaaagactatcataaagctaaa
 gtttacgggttctaaaacattttggtaaaaggtatcgttcagaccactcgatgaatttagtgat

gggttcattaattaaatatacagagatgaaatggctaacgcctgatggccattatattcat
 ggtaaaggaattagaccagatgtagtatctcaacacaaaaatccaatcactcaatgctc
 attccagataacaaaaacttatcatcaaggtgaaaaagataaaaaatgttaaaacgatgaaa
 ataggtctaaaagcttttaggttatccaattgataacgaaacaaacatatattgacgaacaa
 5 ttagaatctgctattaaaaacatttcaacaagacaataatttaaaagttaatggcaatttt
 gataaaaaacaaatgataaatttactgaaaaactagttgaaaaagcgaataaaaaagat
 actgttttaaacgatttactaaacaaactaaaataa

Sequence 406

10 MNDHQKNHATSQDDNTMSTPSKNSKHIIKLWHFILVILGIILLTSIITVVSTILISHQK
 SGLNKEQRANLKKIEYVYQTLNKDYYKKQSSDKLTQSAIDGMVKELKDPYSEYMTAEETK
 QFNEGVSGDFVGIGAEMQKKNEQISVTSMPMKDSPAEKAGIQPKDIVTQVNHHSVVGKPLD
 QVVKMVRGKKGTYYVTLTIKRGSQEKDIKIKRDTIHVKSVEYEKKGNVGVLTINKFQSNST
 15 GELKSAIIKAHKQGIIRHIIIDLRNPNPGLLDEAVKMANIFIDKGNTVVQLEKGDKEELK
 TSNQALKQAKDMKVSILVNEGSASASEVFTGAMKDYHKAKVYGSKTFGKGIVQTTREFS
 GSLIKYTEMKWLTDPGHYIHGKGIRPDVSI STPKYQSLNVI PDNKTYHQGEKDKNVKTMK
 IGLKALGYPIDNETNIFDEQLESAIKTFQQDNNLKVNGNFDKKTNDKFTEKLVKANKKD
 TVLNDLLNLK*

20 Sequence 407

Contig_0475_pos_2540_2037,

putative peptide of unknown function

atgattagattagcaactaaagatgatttacttagtattactcaattagtc aaagaggct
 aaacagatttatggaagaattcaacaacaaccaatgggatgatgaatatcccgcgaaagag
 25 cattttgaagaagacatcgaaaaataaaacactatatgttttagacggtgatcatacaatt
 tatgggtttattgttatcgacccaaatcaatcgagtggtatgatgacattgattggcct
 gttaatcgaaatggggcatacgttattcacagattagctggatcaaaacaataaaaggt
 gctgcgactgaacttttccaatttgccattgacttagcaaatgaacatgatattcatgctc
 attttaacagatacatttgccctcaataaaacctgctcaaggattatttgaaaagtttggt
 30 ttactaaagttgatgagatagagatagattatcatccttttgatagaggggcacctttt
 tatgcatattataaaaaacataataa

Sequence 408

35 MIRLATKDDLLSITQLVKEAKQIMEEFNNQWDDEYPAKEHFEEDIENKTLVLDVDHTI
 YGFIVIDQNQSEWYDDIDWPVNRNGAYVIHRLAGSKQYKGAATELFQFAIDLANEHDIIH
 ILTDTFALNKPAQGLFEKFGFTKVDEIEIDYHPFDRGAPFYAYYKNI*

Sequence 409

Contig_0475_pos_2019_946,

40 is similar to (with p-value 1.0e-63)

>gp:gp|AF068902|AF068902_4 Streptococcus pneumoniae D-glutam
 ic acid adding enzyme MurD (murD), undecaprenyl-PP-MurNAc-pe
 ntapeptide-UDPGlcNAc GlcNAc transferase (murG), cell divisio
 n protein DivIB (divIB), orotidine-5'-decarboxylase PyrF (py
 45 rF), and orotate phosphoribosyltransferase PyrE (pyrE) genes
 , complete cds; and unknown genes. NID: g4009477.

atgacaaaaattgcatatacaggtggaggaacagtaggacacgtttcagtgaatttaagt
 ttaattcctacttgcattgaaaaaggacacgaagcattttatattggttcaaaacatggt
 attgaaagggaatgatagagtcacaaactccctgatattcaatattatccaatatcaagc
 50 ggtaaattacgtcggttatctatcttttgaaaatgcaaaagatgtctttaaagttttgaaa
 ggaatttttagatgcacgtaaaaatacttaaaaaacaaaaaccagacttacttttttcaaaa
 ggtggttttggttaggttcggttagttatagccgcacgttcttttaaaattccaactatc
 atacacgaatcagatttaactcctggattagctaataaaatttcttttaaaatttgctaag
 aaaatatacacacactttgaagatacacttacatatcttccaaaagataaagctgatttt
 55 gtggggctactgtacgtgaggacttaaaacaagggaataaagaaagaggatatcaactc
 actgattttgataaaaaataaaaaagtggttattagtcattggggagggaagtttaggtagtaaa
 aaacttaataatatcattcgtaaaaatattgaggcacttctccacgattatcaaattata
 cacttaactggaaaaggacttggtgatgactcaatcaataaaaaaggttatgttcaattt
 gaatttggttaaagacgacttaactgatttattagcaatcactgatactgttgtaagtcgt

gcagggttctaacgcaatttatgaatttttaacgctacgtataccgatgttactcatcccc
 ttaggacttgatcaatcaagaggagatcaaattgataatgctaaaaactttgaatctaag
 ggttatggtcgtcatattcctgaagatcaacttacagaagttaacttattgcaagaatta
 aatgatattgaattacatcgtgaatctattattaacaaatggaaacatatcaagagagt
 5 tacacgaaagaagatttatttgataaaaattattcatgatgcattaaacaagtag

Sequence 410

MTKIAYTGGGTGVGHVSVNLSLIPTISIEKGHEAFYIGSKHGIEREMIESQLPDIQYYPIS
 GKLRRLYSFENAKDVFKVLKGILDARKILKKQKPDLLFSKGGFVSVPVVIAARSLKIPTI
 10 IHESDLTPGLANKISLKFACKIYTTFEDTLTYLPKDKADFGATVREDLKQGNKERGYQL
 TDFDKNKKVLLVMGSSLGSKLNNIIRQNIALLHDYQIIHLTGKGLVDDSIKKGYVQF
 EFVKDDLTDLLAITDTVVSRAGSNAIYEFLTLRIPMLLIPLGLDQSRGDQIDNAKNFESK
 GYGRHIPEDQLTEVNLLQELNDIELHRESIIQMETYQESYTKEDLFDKIIHDALNK*

15 Sequence 411

Contig_0475_pos_933_319,

is similar to (with p-value 3.0e-21)

>pir:pir|S32217|S32217 hypothetical protein 2 - Bacillus meg
 aterium >gp:gp|Z21972|BMCTP450A_3 B.megaterium cytochrome P4
 20 50meg, ORF1 and ORF2 genes. NID: g288298.
 atgaatcgatggaaacgcatttcattgcttattgtttttacacttatttttggataata
 gctttttttcatgaatcaaggcttggaatggatagataacgaagtatatgaatttatt
 tattcatctgaaagtttcattaccacatctattatggttaggtgtaacaaaaattgggtaa
 gtttgggcaatggttgcgctatccttattatttagttgcttaccttatgctaaaacgcttc
 25 aagattgagacattattctttgtaataagtaataagcttatctagtacactcaatccacta
 ttaagaatatctttgatagggaaacgtccaacattattgcgtttaattgacatttcaggc
 ttagtttttccaagcggctcatgctatgggctcaacttcattctttggaagcgctatatat
 gtaataaaccgctcatgattcgggtatctctaaaggcgtgtaaatcggtttatgcgcactt
 ttcattttattaatatcaacttctagagtgtatctaggcggttcattaccctacagatatt
 30 attgccggcattattgggtggtgtatctgccttttactcagtagctttattactacctaaa
 cagttaataagcttag

35 Sequence 412

MNRWKRISLLIVFTLIFGIIAFFHESRLGKWIDNEVYEFYSSSESFITTSIMLGVTIKIGE
 VWAMVALSLLLVAIYMLKRFKIETLFFVIVMSLSSTLNPLLNIFDRERPTLLRLIDISG
 FSFPSGHAMGSTSFFGSAIYVINRHDSGISKGVLLIGLCALFILLISTSRVYLGVHYPTDI
 IAGIIGGVFCLLLSTLLLKPQLIA*

40

Sequence 413

Contig_0476_pos_1619_2608,

is similar to (with p-value 2.0e-56)

>sp:sp|P54948|YXEI_BACSU HYPOTHETICAL 37.2 KD PROTEIN IN IDH
 45 -DEOR INTERGENIC REGION. >gp:gp|Z99124|BSUB0021_58 Bacillus
 subtilis complete genome (section 21 of 21): from 3999281 to
 4214814. NID: g2636442. >gp:gp|D45912|D45912_12 Bacillus su
 btilis genome sequence between the iol and hut operon, parti
 al and complete cds. NID: g1408482.
 50 gtgttatatatgtgtactgccatttctttatatacaaaaacaacgttaccattattagct
 agaacaatggactttgcatttgaatttaattggtatcccaaccattgttccacgccattat
 cactaccaatttgatctagattcagacatgcgtcttgaatatggttttgttggaacaaat
 ttaaaagtaggacgtttatagatttgggtgatggtataaacgaaaagggttagctatttcg
 aaccattacttcactggtgaagcctcatcacgtaccataaacggttatggttattttaac
 55 ttagcacctgaggagtttattgtttgggttttaggttttaataaaagtattagcgaatta
 aaacaaaagggttaagaaaatcaatattatgaatgaaaaaatacgaactttgaatatcggt
 cctcctttacatttcatgggtcactgatgaaacaggacataccgtagccatagaacctcac
 aatggcttattaatagttaaagataattatgttcataaccttaacaaatgaacctaaatta
 gattggcatctatctaacttaagaaattacgcttttttaacgccacagaaatcaaccaat

caattaataggtaaagtgcctagtaagatcaatgggctgtgaagcaggaacaaatggctta
 ccgggtggttatacgtcaacagatcgttttatcgcgctacatattaagacaccaacta
 cgctgttcccataatgaagatgaaaatttaataatgaattgttttaaagttctagaatcagtc
 agtatccctcaaggtgcagttatcgatgccataaaatacattacacacaatatcaatta
 5 gtgatggaaagtaaaagaaggttattatattaagccttacttttagcaatcaaattttc
 aaaataaaattaactgaagaccttttaagtaagaatgagatgacattcttacctattaat
 cacgaattaaagataacatcaatacaatag

Sequence 414

10 VLYMCTAISLYTKQRYHYLARTMDFAFEFNGIPTIVPRHYHYQFDLSDMRLEYGFVGTN
 LKVGGRYRFGDGINEKGLAISNHYFTGEASYSTHKRYGYFNLAPEEFIVWVLGFNKSISEL
 KQKVKKINIMNEKNTTLNIVPPLHFMVTDGTHTVAIEPHNGLLIVKDNVHTLTNEPKL
 DWHLNLRNYAFLTPQKSTNQLIGKVLVRSMGCEAGTNGLPGGYTSTDRFIRATYLRHQL
 RCSHNEDENLMNCFKVLSESVIPQGAVIDANKIHYTQYQLVMESKERSYYIKPYFSNQIF
 15 KIKLTEDLLSKNEMTFLPINHELKITSIQ*

Sequence 415

Contig_0476_pos_2841_3542,
 is similar to (with p-value 3.0e-41)
 20 >sp:sp|P39610|THID_BACSU PHOSPHOMETHYLPYRIMIDINE KINASE (EC
 2.7.4.7) (HMP-PHOSPHATE KINASE) (HMP-P KINASE). >pir:pir|S39
 707|S39707 hypothetical protein - Bacillus subtilis >gp:gp|X
 73124|BSGENR_53 B.subtilis genomic region (325 to 333). NID:
 g413923. >gp:gp|Z99123|BSUB0020 97 Bacillus subtilis comple
 25 te genome (section 20 of 21): from 3798401 to 4010550. NID:
 g2636240.
 atggataaagaaacatggtcccatgatgtaaacacctattgatatgaatgttttcgaaaaa
 caacttgaaactgcaatatcaattggacctgatgtattaaaacaggaatgttagggaca
 caagacattattaaacgtgccggagatgttttggatgaatctggtgcagactattttgta
 30 gttgatccagtaatggtttgtaaaggagaagacgaagtacttaaccaggaacacagaa
 gcaatgattcaatatttactacctaagctacagttgttaccgccgaatttattcgaagca
 ggtcaactctctggttttaggaaaattaacatcaattgaggatatgaaaaaagctgctcaa
 gtgattttatgacaaaggcacacctcatgtcattattaaaggtggtaaagcactcgatcaa
 gataaatcttatgacttgtactatgatggccaacaattttatcaattaactactgacatg
 35 ttccaacaaagttataatcatggtgcaggatgcacatttgcgtgctgccacaacagcttat
 cttgcgaacggttaaatctccaaaagaagcaatcattgctgctaaagcattttagcttca
 gcaatcaaaaatggttgaaaatgaatgactttgtaggacctgttgatcatggtgcatat
 aaccgtattgaacagattaacgttgaagtcactgaggtttaa

Sequence 416

40 MDKETWSHDVTPIDMNVFEKQLETAISIGPDAIKTGMLGTQDI IKRAGDVFVESGADYFV
 VDPVMVCKGEDEVLPNGNTEAMIQYLLPKATVVTPNLFAGQLSGLGKLTSEDMMKAAQ
 VIYDKGTPHVI IKGGKALDQDKSYDLYDGQQFYQLTTDMFQQSYNHGAGCTFAAATTAY
 LANGKSPKEAIIAAKAFVASAIKNGWKMNDVFGPVDHGAYNRIEQINVEVTEV*

Sequence 417

Contig_0476_pos_10083_11204,
 is similar to (with p-value 0.0e+00)
 50 >sp:sp|P05425|ATKB_ENTFA POTASSIUM/COPPER-TRANSPORTING ATPAS
 E B (EC 3.6.1.36). >pir:pir|B45995|B45995 Cu2+-transporting
 ATPase (EC 3.6.1.-) - Enterococcus hirae >gp:gp|L13292|ENECO
 PPUMP_2 Enterococcus hirae ATPase (copA) gene, complete cds;
 ATPase (copB) gene, complete cds. NID: g290641.
 atgctcattcaaaatgatgttgattttgattagaaacgtcttgtaactgtgttagtcatt
 55 gcttgctccacatgcttttaggcttggaataaccttttagtcactgcacgttctacttcaatt
 ggtgcacataatggtttaattattaaaaatagagagctctgtagaaatagctcaacatatac
 gattatgtaatgatggataaaactggtactttaactgagggttaacttttctgtgaatcat
 tatgagagctttaaaaatgatttgagtaatgatacaatattaagccttttcgcctcatta
 gaaagtcaatctaatacaccattagctataagattgttgattttgcgaaaagtaaaaat

gtttcatttactaatccacaagacgttaataatattccagggtgtcggattagaagggtcta
 attgataataaaacatatataaaataacaaatgtctcttatcttgataaacataaaacttaatt
 tatgacgatgacttgtttactaaattagctcaacaaggtaattcaatcagttatattaatt
 5 gaggatcaacaagtcatttggcatgattgctcaaggagatcaaattaaagaaagctcaaaa
 caaatgatagctgatttactatcaagaaatattacaccagtcagcttacagggtgacaat
 aatgaagtggcacacgctgtcgcaaaagaattaggtattagtgatgttcacgcacaaactc
 atgccagaagataaggaaagcattataaaagattatcaaagtacgggtaataaagtcag
 atggctcgagacgggtatcaacgatgcgccgagtcctataagagccgatattggtatagca
 attggtgcaggcacagatgttgcagtggaattcagggtgatatacacttgttaaaagtaatt
 10 ccatcagatatcattcatttcttgactctttcaataatactatgagaaaaatgggtgcaa
 aacttatgggtggggtgcagggtataatattgttgctgtacctttagcagctggcgcat
 gcttttatcgggttaataattatcaccagctgtaggagcaatattaatgtctttaagtaca
 gttatagtagcgattaatgcttttacattaaaattaaaataa

15 Sequence 418

MLIQNDVDFALERLVTVLVIACPHALGLAIPLVTARSTSIGAHNGLIIKNRESVEIAQHI
 DYVMDKGTGLTEGNFSVNHYESFKNDLSNDTILSLFASLESQSNHPLAISIVDFAKSKN
 VSFTNPQDVNNIPGVGLEGLIDNKTYKITNVSYLDKHKLNYYDDDLFTKLAQQGNSISYLI
 EDQQVIGMIAQGDQIKESSKQMIADLLSRNITPVMLTGDNNEVAHAVAKELGISDVHAQL
 20 MPEDKESI IKDYQSDGNKVMVGDGINDAPSLIRADIGIAIGAGTDVAVDSDGIILVKS
 PSDIIHFLTLNNTMRKMVQNLWWGAGYNIVAVPLAAGALAFIGLILSPAVGAILMSLST
 VIVAINAFTLKLK*

Sequence 419

25 Contig_0476_pos_9841_7541,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF007865|AF007865_3 Bacillus licheniformis bacitracin
 synthetase operon including bacitracin synthetase 1 (bacA),
 2 (bacB) and 3 (bacC) genes, complete cds. NID: g2982193.
 30 atgataacctgtgcattttatgaagggtggatcgatatacctatcacgatgaatgggaaatta
 gatgtgcgtgcattacctgaaattaatctaaagaataatagaaattatgtagaaccacgt
 aacgatattgaacgcacagtttgcctgattttcgaagagattttacatgttgatcaggta
 ggtgttaaagataaatttcttgaactagggtggacactctcttagagcaacattagttgta
 aaccgtattgaagaaagggttaaaaaaacgtcttaaaagtaggtgatttaatgaaatcgct
 35 actgtagagcaacttggacaacaaattgaagaactgcaaaatgatgtctatgaagtgatt
 cccaaagcaaatgaatcgatcaatatgatttaagtgcgtctcaaaaaagtatgtatctt
 ttatggaaggatcaatcctaaagacacagtgatataacattccattcttatggagattatct
 tctgaacttaattgtatgcaattgcaacgtgcattatctaagttgattgaacgtcatgaa
 atattacgaacacaaatatgtaattgatgacaatgaagttaaacaacgtattgcgacacat
 40 ttttcgctgattttgaagaggtaacgacatctctaagcaacgagcaagatattattcaa
 tcatttatggaaccgtttgatttagaacaaccaagtcagatgcgagttaaatatatacat
 ggaccacaacaagattattttatttatggatactcatcatagtttaattgatggatgagt
 aacacgattttactatctgatttgaacgctttataccaagataaatcattacctgaactt
 aagcttcagtataaaagattatagtgagtggtggtgcacagagacttatctaacaacgt
 45 cacttttgggttacagcaatttgaaaatcagggttccaatattaaatatgcctacggattat
 cctagaccaagttataaaacaaccaacggtaatatgttgacgtttcattacaatcgtaa
 atcaaacagcaattgaaatcttatgtagaacaacatcaagtgacagactttatgttcttt
 gctagtgcaatcatggtattattgcacaaatatacacgtcaggacgatatcgctattggt
 agtgtaatcagtgcgctactcatcgcgatactgaaaatatgttaggtatgtttgctaatt
 50 acacttgatatatcggtgcgaccacatgatcaaaagacatgggatcaattgatggctgag
 atgaaagagatgtgtctaggggcatatgaacatcaagaatatccttttgaaagcttagtc
 aatgatcttgttgatgaaagagatgcttcacataatccgttatttgatgtgatgctcgta
 ctcaaaaataatgaacaaatcatgcgaatttttgacatagtcatttgacacatatcca
 cctcagtcacaacacagctaaatttgatttgcatttattattgaagaagatcaagatgac
 55 tatgtcgtcaatattgaatataatacagatttatataaacaagagaccattcatcatatt
 gctgaacaacttcaaatgattatttaaacatgtaatatctaccgaaaacctaataattcaa
 gatattgatgaaaatgatgacttattaatttgggttgacaagcatgtgaatgattgttct
 ttagacttgccaaaaataagtcataacagcaactttacatgatgtcatgaaagcgaaa
 cgagatgatgtagcacttaaaatgaatggacaatcgatgacgtatcaagaacttgatgat

tattctaatagtatggctcaaacattgatacaaaatggcattcaaaaaggggaacgtgta
 gcccttttaactgaacgaagttttgaaatggttgctagtagtattgctgtattaaaagtt
 ggaggttcttatgtacctattgacgtcacttatcccgataaacgcattgaatttattatt
 5 gaagacgctgaagtcgcagcagtgctcacatatggaaaagcaatatcctcacatatacca
 gtaattaaaattgaagatattgataacactgaaaataataaaaaggttaaatatagaatat
 gcaggaatttggagatgatattgatcatatttatacatctggaacaacaggaaagcct
 aaagcagtatcagtgaaacaacgtaatatattaaatttagtatgtgcttgacaaaaaga
 ctcaatttatccgatgatgaagtcctatctgcagtagcctaattatgtgttcgatgcttcg
 gcaactgatttctactgttag

10

Sequence 420

MIPVHEMKVDRIPIITMNGKLDVRLPEINLKNRNYVEPRNDIERTVCRIFFEEILHVDQV
 GVKDNFFELGGHSLRATLVVNRIEERLKKRLKVGDLKMSPTVEQLGQIEELQNDVYEVI
 PKANESYQYDLSASQKSMYLLWKVNP KD TVYNI PFLWRLSSELNVMQLQRALSKLIERHE
 15 ILRTQYVIDDNEVKQRIATHVSPDFEEVTTSLTNEQDIIQSFMEPF DLEQPSQMRVKYIH
 GPQQDYLFMDTHHSINDGMSNTILLSDLNALYQDKSLPELKLQYKDYSEWMVHRDLKQR
 HFWLQQFENQVPILNMPTDYPRPSIKTTNGNMLTFHYNRQIKQQLKSYVEQHQTDFMFF
 ASAIMVLLHKYTRQDDIAIGSVISARTHRDTENMLGMFANTLVYRGRPHDQKTWDQLMAE
 MKEMCLGAYEHQEYPFESLVNDLVDERDASHNPLFDVMLVLQNNETNHANFGHSQ LTHIP
 20 PQSTTAKFDLSFIIIEEDQDDYVVNIEYNTDLYKQETIHIIAEQLQMI IKHVISTENLKIQ
 DIDENDLLIWLDKHVNDCSLDLPKNKSIQQLLHDVMKAKADDVALKMNGQSMYQELDD
 YSNSMAQTLIQNGIQKGERVALLTERS FEMVASMI AVLKVGGSYVPIDVTY PDKRIEFII
 EDAEVA AVLTYGKAISSHIPVIKIEDIDNTENNKR L NIEYAGNLEDDMYHIYTS GTTGK P
 KAVSVKQRN ILNLVCAWTKRLNLS DDEVY LQYANYVFDASATDFYC*

25

Sequence 421

Contig_0476_pos_7178_6864,
 30 putative peptide of unknown function
 gtgcaacaaaatttagggaaaatggaaagtttattactttctgcaagacattttctatgg
 agtacagctagagggtatcaatcatatacagaggatgcacaaatatggaatgaaacctca
 gcaagtaaaagtgggtgtaataatgaaccaaggtatagaaatcgttgatttagctatgagaata
 gttggagctaagagcttagaaatgagcagacctcttcaacggtactatagagatatacgt
 35 gctggattacataatccaccaatggaagatatggcttacactaatattgctaaaagtatt
 acaacaaaactttaa

Sequence 422

VQQNLGKMESLLLSARHFLWSTARGYQSYTEDAQIWNETSASKVVMNQIEIVDLAMRI
 40 VGAKSLEMSRPLQRYRDIRAGLHNPPMEDMAYTNI AKSITNKL*

Sequence 423

Contig_0476_pos_2968_2609,
 is similar to (with p-value 2.0e-20)
 45 >sp:sp|P39610|THID_BACSU PHOSPHOMETHYLPYRIMIDINE KINASE (EC
 2.7.4.7) (HMP-PHOSPHATE KINASE) (HMP-P KINASE). >pir:pir|S39
 707|S39707 hypothetical protein - Bacillus subtilis >gp:gp|X
 73124|BSGENR_53 B.subtilis genomic region (325 to 333). NID:
 g413923. >gp:gp|Z99123|BSUB0020_97 Bacillus subtilis comple
 50 te genome (section 20 of 21): from 3798401 to 4010550. NID:
 g2636240.
 atgtcttgtgtccctaacattcctgttttaatagcatcaggtccaattgatattgcagtt
 tcaagttgttttttcgaaaacattcatatcaataggtgttacatcatgggaccatgtttct
 ttatccattgtttacaatagatgttaaagcgaccattccatatacatcaagttcttggaa
 55 gttttaagatctgcttgcataccggcaccagcacttgtatctgaaccagctatcgtaaat
 acttttttttaaagccatcattcattcactccattaatttctagtgtctttatcatatca
 tgtttatcgcgtagcgtaaattattataatttttaaagtcaaatcaatcatcatacttag

Sequence 424

MSCVPNIPVLIASGPIDIAVSSCFSKTFISIGVTSWDHVSLSIVTIDVKATIPYTSSSWN
VLRSAIPAPALVSEPAIVNTFFKAIHSLPLISSVFIISCLSRTLNYYNFKMQINHHT*

5

Sequence 425

Contig_0476_pos_1461_811,
is similar to (with p-value 3.0e-64)

>pir:pir|S39712|S39712 hypothetical protein - Bacillus subti
lis

10

atgaaatgggtcagaggtatttcatgatataacaacgcgccatgattttcaggcgtatgcat
gacttttttagaaaaagaatatacagactcaaaccgtctatccagatatataaaatatctat
caagcattttgattttaacgcgctttgaagatatcaaggttggtatttttagggcaagatcct
tatcacggctcctaatacagcacatggtttagcattttcagtgcaacctcatgctaaattt
15 ccaccatctttaagaaatatgtatcaagaactagaaaaatgatatagggtgtcatagaact
tcgctcatttacaagactgggcaagagaaggtgtcttgttattaaatacggatttgact
gttcgacaaggtgaagcacattcacatcgaaatattggatgggaaacattcacggatgaa
atcatacaagctgtttctaattatcgtgagcatgttgtttttattctgtgggaagaccg
cctcaacaaaaggaacgatttcattgatacatctaacacttaatacattaaatcgccacat
20 cctagtccactatcggcttttagaggatttttttggttctaacccttattcaactacaaat
aactatttaaaatctaaagggaaaacaccagttcagtggtgtgaaagtttag

Sequence 426

MKWSEVFHDITTRHDFQAMHDFLEKEYTTQTVPDIQNIYQAFDLTPFEDIKVVLGQDP
25 YHGPNOAHGLAFSVQPHAKFPPSLRNMYQELENDIGCHRTSPHLQDWAREGVLLLNTVLT
VRQGEAHSHRNIGWETFTDEIIQAVSNYREHVVFILWGRPAQQKERFIDTSKHLIIKSPH
PSPLSAFRGFFGSKPYSTTNNYLKSKGKTPVQWCES*

Sequence 427

Contig_0476_pos_810_442,
putative peptide of unknown function

gtgagaataatgaacaaagaacagattctacaattgattgagcaagaattgatataagca
gatgaagctcagacagatacggaaatttgaaaagcatatgtatgctatacacatgctcaca
tctcttggttagttctcatcaaagtcgttctacaatagagaaattaaatcattctaaacca
35 atgaatagtaatatcaaagatgattatgagatgaaacaacagtccttcacaaaaacatcat
gtaactgcagctgaaatagaagcaatgggtggtaaagtaccacaatcaatgaaaaagcat
catacttctaataatatgatgattacagatgatcaagttggtaaatggtgaatctattttt
gatttttaa

40

Sequence 428

VRIMNKEQILQELIQADEAQTDTEFEKHYAIHMLTSLVSSHQSRSTIEKLNHNSKP
MNSNIKDDYEMKQSSQKHHVTAAEIEAMGGKVPQSMKKHHTSNNMMITDDQVNGESIF
DF*

45

Sequence 429

Contig_0476_pos_420_49,
is similar to (with p-value 3.0e-27)

>sp:sp|P39619|YWDK_BACSU HYPOTHETICAL 12.0 KD PROTEIN IN UNG
-ROCA INTERGENIC REGION. >pir:pir|S39716|S39716 hypothetical
50 protein - Bacillus subtilis >gp:gp|X73124|BSGENR_62 B.subti
lis genomic region (325 to 333). NID: g413923. >gp:gp|Z99123
|BSUB0020_88 Bacillus subtilis complete genome (section 20 o
f 21): from 3798401 to 4010550. NID: g2636240.

55

atgatgaaagttttattatttttaggtgcatataatgcaatgatggctgtcggtactggc
gcatttggagcacatgggttggaagataaattatcagataaatacatgtcaatatgggaa
aaagcaacaacttatcaaattgtatcatggattaggtctgttagttataggtttaataagt
gggtacaacatcaattaatgtaaattgggctggttggtattattctttggtattgtcttt
ttcagtggttcttctgtatttcttagccttaacacaagttcgtatttttaggtgcaattacg
ccaataggtggtgttctattttataattgggttggttcttctgtgattgctacacttaaa

ttcgctgggtaa

Sequence 430

MMKVFIILGALNAMMAVGTGAFGAHGLEDKLSDKYMSIWEKATTYQMYHGLGLLVIGLIS
5 GTTSINVNWAGWLLFFGIVFFSGSLYFLALTQVRILGAITPIGGVLFIIIGWLVLVIATLK
FAG*

Sequence 431

Contig_0477_pos_802_1206,
10 is similar to (with p-value 8.0e-16)
>sp:sp|P49856|YKKC_BACSU HYPOTHETICAL 11.9 KD PROTEIN IN HMP
3'REGION. >gp:gp|D78189|BAC168TRP2_6 Bacillus subtilis hmp
DNA for 7 ORFs, complete cds. NID: g1063245. >gp:gp|AJ002571
|BSAJ2571_28 Bacillus subtilis 168 56 kb DNA fragment betwee
15 n xlyA and ykoR. NID: g2632001. >gp:gp|Z99110|BSUB0007_191 B
acillus subtilis complete genome (section 7 of 21): from 119
4391 to 1411140. NID: g2633472.
atgtttgtgagaagccattgcctcgaagttacgaatttgagaaaatggttctattacttt
aataggaggaataaaaagaatgcaatggcttaaagttatattagccggttttattgaaatc
20 atctgggtcactggacttgatcaagcgcaactcattgtttacatggatattaccctcttt
tttattgctttaagcttttttctagtcattgatgcttcgaagcacttaccagttggtacg
gtatatgcattttttgtcggaaatcggtgctgttggtacagtgttagttgatatgattttc
ttcaaccaaccatttactttcactaaaatcttttaataatgacccttatttttaggaata
ataggattaaaactgacaactgatgcaacgaaagaaggagataa
25

Sequence 432

MEVRSHCLEVTNLRKWFYFNRRNKRMQWLKVILAGFIEIIWVTGLDQAHSLETFWIFTLF
FIALSFFLVIDASKHLPVGTVYAFFVGIGAVGTVLVDMIFFNQPFFTFKIFLIMTLILGI
30 IGLKLTTDATKEGR*

Sequence 433

Contig_0477_pos_1212_1529,
putative peptide of unknown function
atggcttggttattttctaataatgatagccggaagttttgaaattttgggcgttggtctatta
35 aatgaactatcacgtacaaagaataaaatttatgtcatttttttaggattagcattttata
ttaagtttttagtacattaaaatttgcaatggtatctattcctatgggtactgcatacgc
atatggacaggaatttggtacagctggtggtacattaattggaatgattttttatagagaa
tctacacgtttaagtagaattttatgtattttattaatcatcatttcagttggttgatta
40 cgtttaataagttattaa

Sequence 434

MAWLFLMIAGSFEILGVVLLNELSRTKNKIYVIFLGLAFILSFSTLKFAMVSIPMGTYA
IWTGIGTAGGTLIGMIFYRESTRLSRILCILLIIISVVGLRLISY*

Sequence 435

Contig_0477_pos_1643_2560,
putative peptide of unknown function
gtgggacttaacttattgaaagaacattttgaggtagacatgtatgatggcgaggggctt
attgataaagaaaccttaaaaaagggttagaacatgcagatgcattaattagtttacta
50 tcaacttctgttgataaagatattattgatagtgcataaaccttaaaattatagcgaat
tatgggtgcaggttttaataatattgatgtcgaatatgcaagacaacaaaatatagatgtt
acaaatacaccacacgcttcgacaaatgctactgctgatttaacaatcggtttaatttta
tcagtagcgcgtagaattgtagaaggagatcatttatccagaacaacagggttttgatggt
tgggcacccttattcttcgaggcagagaggtatcaggaaaaactattggtattataggc
55 ttagggtgaaattggagggtgcagtgcataaacgcgcacgcgcatttgatatggatggtctg
tactactggtcctcatcgtaaggaagaaaaagaacgagatatcggtgcgaaatatgtagat
ttagatactttacttaaaaatgcagattttattacaatcaatgcggcatataatccatca
ctgcatcatatgattgatactgaacaatttaataaaatgaaatctactgcctatttaatt
aatgcaggacgtgggtccaatagtaaatgaacaatctttagttgaagcccttgataataaa

gctattgaaggtgctgcattggatgtatatgaatttgagccagaaatcactgatgcatta
 aaatcatttaaaaaacgttggtgcttacacctcacattggtaatgcaacatttgaagctaga
 gatattgatggctaaaaattgttgcgaatgatacaataaaaaaattaaatggtgatgaacct
 cagtttattgtcaattaa

5

Sequence 436

VGLNLLKEHFEVDMYDGEGLIDKETLKKGVEHADALISLLSTSVDKDIIDSANLKIIAN
 YGAGFNNIDVEYARQQNIDVTNTPHASTNATADLTIGLILSVARRIVEGDHLSRTTGFDG
 WAPLFFRGREVSGKTIGIIGLGEIGGAVAKRARAQFMDVLYTGPHRKEEKERDIGAKYVD
 10 LDLLKNADFITINAAYNPSLHHMIDTEQFNKMKSTAYLINAGRGPIVNEQSLVEALDNK
 AIEGAALDVYEFEPITDALKSFKNVLTTPHIGNATFEARDMMAKIVANDTIKKLNGDEP
 QFIVN*

Sequence 437

15 Contig_0477_pos_6327_5923,
 is similar to (with p-value 2.0e-16)
 >sp:sp|Q02115|LYTR_BACSU MEMBRANE-BOUND PROTEIN LYTR. >pir:p
 ir|A47679|A47679 lyt divergon expression attenuator LytR - B
 acillus subtilis >gp:gp|M87645|BACLYTABCD_1 Bacillus subtilis
 20 is membrane bound protein (lytA and lytR); amidase enhancer
 (lytB); and amidase (lytC) genes, complete cds's. NID: g1431
 55. >gp:gp|299122|BSUB0019_62 Bacillus subtilis complete gen
 ome (section 19 of 21): from 3597091 to 3809700. NID: g26360
 29.
 25 atgggggctaatactctttaaaggtgaaaaaacacacgtagatggtgatgctgccatg
 gactttattagaagtcgtaagaagatggggcaggaggcgattttggtagacaagagcgt
 cagcaacttatcttagaagcgatggcagataagatgacaagcgcttcttcaatcactcat
 ttaatacattaatgaatcaaattcagaaaaatgttaaacagatttaaaattaggtgat
 ctaatacaattagaactaagtataaagatgctaatagaccaagttaatcgacalcagtta
 30 gaggggtgaaggtggtatacaaaatgacggtttgtactatttcataccaagtgatgcactc
 aaaaatgaaaatacacaattactaagagacaatttaaaattataa

Sequence 438

35 MGANHFVKGEKTHVDGDAAMDFIRSRKEDGAGGDFGRQERQQLILEAMADKMTSASSITH
 FNTLMNQIQKNVKTDLKLGDLNTRTKYKDANDQVNRHQLEGEGGIQNGLYYFIPSDAS
 KNENTQLLRDNLNL*

Sequence 439

40 Contig_0477_pos_4992_4216,
 is similar to (with p-value 3.0e-45)
 >gp:gp|U71377|SEU71377_4 Staphylococcus epidermidis autolysi
 n AtlE and putative transcriptional regulator AtlR genes, co
 mplete cds. NID: g2267238.
 45 atgttaaagaataactcggcttcgaatgacgacattgttcataattagcataactcgtcatt
 ttagcaatactttttcttataatttgatacaaacctttttaaaatgatgttaaacatata
 ttaaaagaagcggtgtctcttcaaaacaagtgagggaatatccataactaaagaagttaat
 ggtaagtttatatatgcttccaaacaagatatagagaaagctatgcaataaaacatagt
 gataatgatttgaagtacatggatatatcagaaaaagtagctatgtcagagaaggaagtt
 aaccatatcttaaaaggaaaaggtattttagaaaaataagggtatcaacgtttattaaagcc
 50 caagataaataatgaagtgaatatcctatatctcatcagtcagtcactagttgagacagga
 aatggtcaatcagatttatcaaaaggaattaaagaaggttaaccatcactattataacttt
 tttggtattggtgcttttgatgaagatgctgtaaagactggtaagagttttgctaaacag
 aagaagtggaccactcctgaaaaagcgataaatgggtggcgcggtggtttgtgagataccat
 tacttttaaaataatcaattgagcttatatcaaatcggtggaaccacaaaatccaggc
 55 caacatcaatatgctagtgatattcagtgggccaataatatagctgatttaattggagaaa
 tactatgataaataatggaataaaaaaagatcatataagaaaaaattattacaaataa

Sequence 440

MLKNTRLRMTTLFIISILVILAILFLIFDTNLFKNVDKHTFKEAVSLQTSEGNIIHTKEVN

GKFIYASKQDIEKAMQIKHSDNDLKYMDISEKVPMSSEKEVNHILKGKGILENKGSTFIKA
QDKYEVNILYLISHALVETGNGQSDLSKGIKEGNHHYYNFFGIGAFDEDAVKTKGSFAKQ
KKWTTPEKAIMGAWFVRYHYFKNNQLSLYQMRWNPQNPQGHQYASDIQWANNIADLMEK
YYDKYGIKKDHIRKKYYK*

5

Sequence 441

Contig_0477_pos_3832_2708,

is similar to (with p-value 4.0e-24)

>gp:gp|U29897|PAU29897_1 Pseudomonas aeruginosa FAD binding
protein homolog gene, partial cds. NID: g912581.

10 atgaaaatagcaatagtagggcgaggtataggtggtttaactgctgctgcttattagaa
gaacaagggtcatcaagttaaagtggttgaataaaataacttctataaacgaattaaagcgct
gggtattgggataggagataatgttttaaaaaaattagggcatcatgaccttgctaaagggc
attaaaaatgctggtcaaaatcttaccgcaatgaatatttatgatgagcaaggcaccacca
15 ttaatgagcgctaaattgaagtctcattccctaaatgtcgcatctatagacaaacttta
attgagatcatacagtcattatgtcgaagaatcatctattcacacaggatttaaagttact
aaaattgaacaaacgagttgtaaggttacctacattttaccaaacaggaaagtgaatcg
tttgatttgggtgattggtgctgatgggttacattctgtagtaagagagctctgtagggtgca
cgaaactaaaattcgttacaatggttacacatgttttagaggcatggttgaagatgtacaa
20 tttaatgaccaacatggttgcaatgaatattggggtgttaaaggacgagtaggtatagtc
ccattaattaatcaacgtgcttattgggtttattactgttcattgctaaagaaggagatcca
aaatcatcaatcttttgaaaaccccatcttcaagcatatttaatacactttccaaatgaa
gtgagaaatgtgttagaaagacaaagtgaacaggtatattacttcatgacatatatgat
ttaaaccactgaagacattcgtttatggacgtactattttaatggcgatgctgcgcat
25 gccactacgcctaataatgggacaaggtgctagtcaagctatggaagatgcaattgtatta
gtgaattggttagaaaaatatgattttaataaagcgattgagcggttatgataaaacttaga
gttaaacataaccacaaaagtgattagcggttcgaaaaagataggtaaaatggctcaaaag
catcataaattaactgttaaaacttagaaaatccgcgatgaaattaataccaaatgctttg
gcatcagctcaaacaaaattttttatacaaatccaaagaaaagtaa

30

Sequence 442

MKIAIVGAGIGGLTAAALLEEQGHQVKVFEKNTSINELSAGIGIGDNVLKKGHHDLAKG
IKNAGQNLAMNIYDEQGTPLMSAKLKSLSLNLVALSRQTLIEI IQSYVEESSIHTGFKVT
KIEQTSCKVTLHFTKQESSEFDLCIGADGLHSVVRRESVGARTKIRYNGYTCFRGMVEDVQ
35 FNDQHVANEYWGKGRVGVPLINQRAYWFITVHAKEGDPKYQSFGKPHLQAYFNHFPNE
VRNVLERQSETGILLHDIYDLKPLKTFVYGRITILMGDAHATT PNMGGQASQAMEDAIVL
VNCLEKYDFNKAIERYDKLRVKHTTKVIRRSKKIGKMAQKHHKLTVKLRNTAMKLIPNAL
ASAQTKFLYKSKEK*

40

Sequence 443

Contig_0478_pos_5223_6236,

is similar to (with p-value 5.0e-32)

>gp:gp|AL034447|SC7A1_23 Streptomyces coelicolor cosmid 7A1.
NID: g4007715.

45 atgctctcaagagcaccatttggatttaaaggcaatcatatacctgctttaattggctgg
gtagggtcaagttggttggttatctgttaattgtttctacaggaactttaactcttctggct
ttattcaatacttttgggttttaagactagtagacatttctaatgttgatgagtttagcgatt
tttgcctgggctagttattatatctgttctttttcacaaaaagtagtctgtatcagtacaa
acatttttcacatatgtatttgggtgcattaaccttattagttataacaattttaattact
50 aatactgattggaacgcctttttctatgaaatctgggtcttggcttaaaaggttttcta
cctgcattagcctttgtaatagtagggactggattgagttggactaacgcagctgcagat
tatagccggttttcaaaaaaaatcgaaacagttctttatcaataatcactagtggttacagct
ggcgcggtttatcccttttatttctcattataagtagtgaattttattagctacttcagag
gcacaatttagcaaatgcagaaaatccaatattattaattagcgaagtactaccaaatgg
55 atgacagtaatttacttaatatctgcttttaggtggccttactcctatgtgttttttaggt
ttaagtcctcaagattaattatgagtacttttgatttgaaagtaaaaaattctacagtt
attattattcattcaattattatttggcattcctatttatgtcttagtagtttccaga
aattttctcgctttttttgaaatgttttttaggagttttgggtattggattagctgcttgg
tctgcaattttcattgttgattatgcaacattgagaaaaatataggctatgaaaaaaa

ttggtttgcgatccccagtataatagtctgaatattaaaacagtaatggctctggagtata
gcagtaaatagtaggtgcattaataaacattttttattcttcaagtttgatataa

Sequence 444

5 MLSPAPFGFKGNHIALIGWVGQVWLSVNVSTGTLTLLALFNTFGFKTSTFLILMSLAI
FAGLVIIISVLFQKVLVSVQTFFTYVFGALTLLVITILITNTDWNALFSMKSGLKGLF
PALAFVIVGTGLSWTNAADYSRFQKSNSSLSIITSVTAGAFIPLFLIISTGILLATSE
PQLANAENPILLISEVLPNWMTVIYLISALGGLTPMCFGLKSSRLIMSTFDLKVKNSTV
10 IIIHSIIIIAIIPIYVLVSRNFLAFFEMFLGVLGIGLAWSAIFIVDYATLRKNIGYEKK
LVCDPQYNSLNKTMVMSIAVIVGALINIFILQVLI*

Sequence 445

Contig_0478_pos_6370_6675,
putative peptide of unknown function

15 gtgatcggcagtgctgttattgatgtgatttttaaatgttaatagtataaccaagtagtgga
tcagacgaatttgccactctgagaagacaatagtaggtggttgcgcataataatgtaggc
gataacttagtcagttcaaagctaattatgatttgatggtgcccgttgccgatggtcct
aatggaacaattattgaaaataagttaaaaaagaaggcaaaacttcattattaaataat
atattaggtgataatggttgacgttatgtactgagagatcccctcataatttcccaaaa
20 gcgtaa

Sequence 446

VIGSAVIDVILNVNSIPSSGSDEFAHSEKTIVGGCAYNVGDILSQFKANYDLMVPVGDGL
NGTIIENKLKKEGKTSLLNNILGDNGWTLCTERSPHNFKPA*

25

Sequence 447

Contig_0478_pos_8029_8592,
putative peptide of unknown function

atgacttggaataaagaattggttttaagaattgatcttaataagtagtattatattat
30 gtgtcaggttattcttttgaaacctctcagacgaagtttattagaagaatttagtcgt
ttaaacgagaaaactacaattatttttgacccctcaccaaggattaataaaatgaactgt
gagagtataaggaagttgcttgaaataaacacaatagtagacatgccaacgaagtgaaata
ttacaattgagtagtgagaatcatgtgaaagatgcggcattagaagtaagtaaacagact
aatcaacctgtgatagttacattaggcaacaaaggtactcttatagcaaataagtgtaa
35 gtaagatttttagagggggaaaaggttccctgtaactgatactataggcgctggtgattca
cacacagcagcttttatagcaggttgctagataaccaaagtagtattgaaaagcttgata
tggggaaacgaagtagcatctaaaattgtgcaagaacgaggtggaaatacggatatattc
aatcctatagataaagaatattaa

40

Sequence 448

MTWKKNWFKEIDLKNDYIYVSGYSFEPPSDEVLLLEFSRLNEKTTIIFDPSPRINKMNC
ESIRKLEINTIVHANEGEILQLSSENHVKDAALEVSKQTNQPVIVTLGNKGTLIANKCK
VKILEGEKVPVTDITIGAGDSHTAAFIAGLLDNQSIKACIWNNEVASKIVQERGGNTDIF
NPIDKEY*

45

Sequence 449

Contig_0478_pos_10650_0,
is similar to (with p-value 1.0e-69)

>sp:sp|P16468|MAOX_BACST MALATE OXIDOREDUCTASE (NAD) (EC 1.1
50 .1.38) (MALIC ENZYME). >pir:pir|A33307|DEBSXS malate dehydro
genase (oxaloacetate-decarboxylating) (EC 1.1.1.38) - Bacill
us stearothermophilus >gp:gp|M19485|BACMAL_1 B.stearothermop
hilus malic acid gene, complete cds. NID: g143164.
atgtctttaagagatgacgcttttagaaatgcatagagagaaccaaggtaaactagaaatt
55 acaccaaattgttaaagtgacaaataagcaaatgaagcctagcactacacctggcggt
gcagaacctgttaaagaaatccatgaagattcaagaaaagtatatgagtacactattaaa
ggaaatacagttgctgttgtaacagatggaactgctgttctcggttttagggaatattgga
gcagaagcaagtagtccagtaattggaaggaaaggcagcactgttcaaaagttttgcgggt
attaatggtgtgccaatagctctagatacaactgacactcaagaaatcataaaaacagta

aaacttattgcaccaaactatggtggaattaatcttgaagatatatcagctccccgctgt
 tttgaaattgaagaaaccttaaagaaagagaccaatatacctattttcatgacgatcaa
 catggtacagctattgttactatggctgggttaataatgctttaaaaattgtagataaa
 gagttaacggatataaaaagttgtattaaatggcaggtgcagcaggtatcgctatagtg
 5 aagttacttcatgcttatggtgtgaataatatgattattcacaccataagca

Sequence 450

MSLRDDALEMHRENQKLEITPNVKVTNKQQLSLAYSPGVAEPCKEIHEDSRKVYEYTIK
 GNTVAVVTDGTAVLGLGNIGAEASIPVMEGKAALFKSFAGINGVPIALDTTDTQEI IKT
 10 KLIAPNYGGINLEDISAPRCFEIEETLKKETNIPFHDDQHGTAIVTMAGLINALKIVDK
 ELTDIKVVLNGAGAAGIAIVKLLHAYGVNNMIHTISX

15 Sequence 451

Contig_0478_pos_10987_10646,

is similar to (with p-value 1.0e-28)

>gp:gp|U35659|SBU35659_1 Streptococcus bovis malic enzyme ge
 ne, complete cds. NID: g1006838.

20 gtgtcagttgtatctagagctattggcacaccattaatacccgcaaaaacttttgaacagt
 gctgcctttcccttccattactggaatacttgccttctgctccaatattccctaaaccgaga
 acagcagttccatctgttacaacagcaactgtatcttcttaatagtgtactcatatact
 tttcttgaatcttcatggatttctttacaaggttctgcaacgccaggtgagtagctagg
 ctttaattgttgcttatttgtcactttaacatttgggtgtaatttctagtttaccttggttc
 25 tctctatgcatttctaaagcgtcatctcttaaagacatttaa

Sequence 452

VSVVSRAIGTPLIPAKLLNSA AFPSITGILASAPIFPKPRTAVPSVTTATVFPLIVYSYT
 30 FLESSWISLQGSATPGEYARLNCCLFVTLTFGVISLPLWFSLCISKASSLKDI*

Sequence 453

Contig_0478_pos_9610_9278,

putative peptide of unknown function

35 gtgtcagtttcttataaaattgctaaaaatctattggatcacatgtacaaaaatgaggat
 agatttcttagcattacatagaaactacgaaaaggaaaaactattatttcttactttacct
 attattggactcataactataataggaagttcatttctcttcgattatttaatatatttaa
 ctgaataatacgtctgtagaaatattaggggtccattcctactgttatatatcaaattatt
 atttgtttttattcatttatgtttcacggctatgttttaataatatttatttataccatt
 40 tggttttttatatatggaaaagtttacaaaataa

Sequence 454

VSVSYKIAKNLLDHYKNE DRFLALHRNYEKEKLLF LTLPIIGLITIIGSSFLFDYLIFK
 LNNTSVEILGSIPTVIYQIIICFIQFMFTAMFLIIFIYTIWFFIYGKFTK*

45 Sequence 455

Contig_0478_pos_4758_3190,

is similar to (with p-value 0.0e+00)

>gp:gp|Z99111|BSUB0008_149 Bacillus subtilis complete genome
 (section 8 of 21): from 1394791 to 1603020. NID: g2633699.

50 >gp:gp|Z97025|BSZ97025_8 Bacillus subtilis nprE, yla[A,B,C,D
 ,E,F,G,H,I,J,K,L,M,N,O] and pycA genes. NID: g2224758.

atggttgacggtgtcgtactagtgttgacgcatatgaaggtacaatgcctcaaactcgt
 tttgttcttataaaaagctttagaacaaaacttaaaaccggttgtagttgtgaataaaatt
 gataaaccagctgctagacctgagggagttgtagatgaagtattagacttattcattgaa
 55 ttggaagcgaatgatgagcaattagacttccagttgtttatgcttcagctgtgaatgga
 acagcaagtttagactctgaaaagcaagacgaaaatatgcaatccctatacagagacgatt
 attgactatgttccggcaccagtagataattcagatgaaccattacaattccaaattgct
 ttactagattataatgattatgtaggtcgtataggcgttgacggtgtgttcagaggtaaa
 atgcgtgtaggtgataatgtatcactaattaaattagatggtacagtttaagaactttcgt

gtgacgaaaatatttgggtactttgggtcttaaacgtgaagaaattgaagaagcacaagca
 ggagacttaatagctgtttcaggtatggaagatattaacgttgggtgaaacagttacacca
 catgatcatcgtagaccattaccggtgttacgtattgatgaaccaaccctagaaatgact
 tttaaagtaaataactctccgtttgctggacgtgaagggtgattatgtaacagctcgacaa
 5 attcaagaaagatttagatcaacaacttgaaacagatgtttctttaaaagttacacctact
 gatcaaccagattcatgggttgttggctggtggaactacacttgctattcttatt
 gaaaacatgagacgtgaaggctttgaattacaggtttctaaacctcaagttattttaaga
 gaaatcgatgggtgtgttaagtgaaccatttgagcgtgtacaatgtgaagtccttctgaa
 aatgccggggcagtgattgagtcattaggtgcacgaaaaggtgaaatgttagatatgatg
 10 acgaccgacaatgggttgacgcgtttaatctttatggtacctgcacgcggtatgattgg
 tatactactgaatttatgtctatgacacgaggttatggaattattaaccatacatttgaa
 gaatttagacctcggttaaagctcaaactcggtggttagacgtaacgggtgcattgatttct
 atggaccaaggtcaagcaacatcttatgagattattaacttagaagatcggtggtgtaac
 ttatggaaccaggtactgaagtatatgaaggatgattggtggtgaacataaccgtgag
 15 aacgatttaacagtaaatattactaaagcaaaagcatcaaaacaaacgtacgttcagctact
 aaagatcaaacacaaacgatgaatcgctcctagaattttaacattagaagaagcggttaca
 ttatcaatgatgatgaattggtggaagtaactcctgaaagtattcgcttaagaaagaaa
 atacttaataaatctgcccgtagaaaagaagcaaaaagagttaaacaattaatgcaagac
 gaacaataa

20 Sequence 456
 MVDGVVLVVDAYEGTMPQTRFVLKKALEQNLKPVVVVNKIDKPAARPEGVVDEVLDLFIE
 LEANDEQLDFPVVYASAVNGTASLDSEKQDENMQSLYETIIDYVPAPVDNSDEPLQFQIA
 LLDYNDYVGRIGVGRVFRGKMRVGDVNSLIKLDGTVKNFRVTKIFGYFGLKREEIEEAQA
 25 GDLIASVGMEDINVGETVTPHHRDPLVLRIDEPTLEMTFKVNNSPFAGREGDYVTARQ
 IQERLDQQLTDSVLSKVTPTDQPSWVAVARGELHLSILIENMRREGFELQVSKPQVILR
 EIDGVLSEPFERFVQCEVPSENAGAVIESLGARKGEMLDMMTTDNGLTRLI FMVPARGMIG
 YTTEFMSMTRGYGINHTFEEFRPRVKAQIGRRNGALISMDQQQATSYAIINLEDRGVN
 FMEPGTEVYEGMIVGEHNRENDLTVNITKAKHQTNVRSATKDQTQTMNRPRILTLEEALQ
 30 FINDDELVEVTPESIRLRKKILNKSAREKEAKRVKQLMQDEQ*

Sequence 457
 Contig_0478_pos_2736_1723,
 putative peptide of unknown function
 35 atggaacgattttgttgtgtaaatcaaatataactatattcaaatgaatccggttagaagcc
 aaattttaaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagcct
 gcttgttttaggaccgacgcttaaacaaacagacagcttacctatacatgagttaatattc
 tttgaattaagagaacgcgtccgttttcatctagaaatcgagaatgaacaaaatcgactt
 aaatttcagatccttgaattactccatcaaacttccctgggttagaagattgtttagt
 40 agtcgatattcaatcattgcactcaacatcgagaaatctttactcattcagacatggtt
 cttgatatcgacaaggaggtactgattacacatatattcaattctacagataaggggaatg
 tcaatggataaagctacaaaatattgcacttcaattaaggggtgattgctcaagaaagctat
 cctaattgtcgatagacattcctttctagtcgaaaaattacgcttacttattcaacaatta
 aaacaatctattcatcatctcaacaattagatgatgccatgattcaattagcacaacaa
 45 ctcgattattttgaaaatattcatcgatacctggtatttggttaagctaagcacagctatg
 attattggggagattggtgatattaagcgatttaaatcaataaacaactcaatgctttt
 gttggcattgatatacaacgatataatcaggtcatacacactgtagagataccatcaac
 aagcggtgtaataaaaaagcgagaaaacttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgatgattattactacaaactaagaaagcagcct
 50 aatgagaaacctcataagactgccatcattgcttgataaatcgattattaaaaacaatt
 cattatcttgtaaatgaatcataaattgtacgattatcaaatgtcaccacattag

Sequence 458
 MERFCCVNQINYIQMNPLEAKFKTSALRSWKTDQADAHKLACLGPTLKQDLSLPIHELIF
 55 FELRERVRFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSIIALNIAEIFTHSDMV
 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLRLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFKSNKQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLFWIMNII RGQHHYDNHVVDYKYLRKQP
 NEKPHKTAIIACINRLKTIHYLVMNHKLYDYQMSPH*

Sequence 459

Contig_0479_pos_2395_989,

is similar to (with p-value 0.0e+00)

- 5 >pir:pir|S19723|S19723 dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Staphylococcus aureus >gp:gp|X58434|SAPDHDNA_3 S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase. NID: g48871.
- 10 atggtagttggagatttcccaattgaaacagatactattgtaataggagcaggtccaggtg
gatatgtcgagccattcgcgcggtcaattaggacaaaaggtaacaatcgttgagaaa
ggtaatttaggtggtgtatgcttaaacgttggtgtataccttcaaaagcattactacat
gcttctcatcgctttgtgaagcgcaaaattcagaaaacttaggggtaattgctgaaagc
gtttcggttaaaactatcaaaaagttcaagaattcaagacttctgtagttaataaattaact
15 ggcggtgttgaggacttttaaaaggtaacaaagtagagattgttagaggtgaagcttat
ttcgttgataacaatagtttacgtgtcatggacgaaaagagtgtctaaacttacaatttc
aaacatgagattatagctacaggttcaagaccaattgaaattccaaattttgaatttggt
aaacgtgttatcgattcaacaggagcttttaattctacaagaagtacctaacaaactagtt
gtagttggtggcgatatacggttctgaattaggtagtctttgcaaactttggctct
20 gaagtactatccttgaaggtgcaaaagatatattaggcggtttgaaaagcaaatgaca
caacctgttaaaaaaggtatgaaagaaaaaggtatcgaaatcgttactgaagcaatggca
aaatctgcagaagaaactgaaaatggtgtcaaagtaacttatgaggcaaaaggtgaggaa
caaactatcgaagctgattatgtattagttacagttggccgtcgccctaatactgatgaa
ttaggattagaagaacttggtctgaaatttgctgatcgtgattactagaagtggaacaaa
25 caaagtcgtacttctattgaaaatatctttgcgattggagatattgtacctggattacca
ttagctcacaaagctagttatgaaggtaaagttgctgctgaagcgatagatggtcaagcc
gcagaggtagactatattggtatgccagcagtttgccttacagaaccagaattagcacia
gttggttatactgaagctcaagcaaaagaagaaggtttatcaattaaagcttctaaattc
ccttatgcagctaattggacgagctttatcattagatgatacaaatggtttgttaagtta
30 attacacttaagaagatgatacgcttattggagcacaagttgtaggtactggcgcatct
gatattatctctgaattaggttttagctattgagtcaggtatgaatgctgaagatatcgca
ttaactgtacatgcacaccaactttaggtgaaatgacaatggaagctgctgaaaaagca
attggttatccaattcatactatgtaa

35 Sequence 460

- MVVGDFPIETDTIVIGAGPGGYVAAIRAAQLGQKVTIVEKGNLGGVCLNVGCIPSKALLH
ASHRFVEAQNSENLGVIAESVSLNYQKVQEFKTSVVKNTGGVEGLLKGNKVEIVRGEAY
FVDNNSLRVMDEKSAQTYNFKHAIATGSRPIEIPNFEFGKRVIDSTGALNLQEVNKLIV
VVGGGYIGSELGTAFANFGSEVTILEGAKDILGGFEKQMTQPVKKGMKEKGIEIVTEAMA
40 KSAEETENGVKVITYEAKGEEQTIADYVLTVGRRPNTDELGLEELGLKFADRGLLEVDK
QSRTSIENIFAIGDIVPLGLPLAHKASYEGKVAAEAIDGQAAEVDYIGMPAVCFTEPELAQ
VGYTEAQAKEEGLSIKASKFPYAANGRALSLDDTNGFVKLITLKEDDTLIGAQVVGTTGAS
DIISELGLAIESGMNAEDIALTVHAHPTLGEMTMEAAEKAIGYPIHTM*

45 Sequence 461

Contig_0480_pos_567_1610,

is similar to (with p-value 0.0e+00)

- >gp:gp|AJ005352|SAA005352_4 Staphylococcus aureus, Sst putative iron transport operon. NID: g3724154.
- 50 atgaaaaaacagctcttatttttattattgtctctagtttttagttttaacggttgtagt
aatagttcgaataataattcaacttcgaaaaagaaaaatagtgattctaaagaaactgta
accatcaaaaatagttttgaagcaagtggttaagaaaaataatggcagtgataagaaaaaa
atctctaataactgtcgaagtaccaaagaatcctaaaaaatgccgttgatttagattatgga
gcgcttgatgtgttgaaagaattaggtgtggctgataaagtaaaaggtttacctaaggt
55 gaaaataaccaacttttacctaatttttagatgaatttaagatgataagtatattaat
actggaaattttaaagaagtgaaactttgataaagttgcatcagctaaaccagatgtgatt
tttatttcaggaagaacagctaatacagaaaaatttagatgaatttaaaaaagctgcacca
aaagctaaagttgtatatgtaggtacaagtgatgacaacttaattaaagatatgaaaaaa
aatacagaaaatttagggaaaatctacgataaagaagataaagctaaaaaaattaataaa

gatttagatagaaaaatatctgatatgaaagataaaaactaaagactttaataagaaagta
 atgtattttattggttaacgaaggtgaactatcaacgtttggaccaggaggaagatttggt
 ggtttagtggttgatacattaggtttaaacctgcagacaaaaaggttagcaaaagcccg
 catggtcaaaatataaataatgaatatattaacaagcagaatccagatgttatttttagct
 5 atggatcgtggttcagttgtaggtggttaaagcaacaacaaatcaagttttaaaaaactaa
 gttataaaaaatgtaaaagcagtaaaaagtaatcatattttacgaattagatccaaaacta
 tggattttctcttcaggatcttcaacgacaactatcaaaacaaattgatgaattaaatgaa
 gtagtagagaaaagttgaaaaataa

10 Sequence 462
 MKKTVLFLLLSLVLTACSNSSNNNSTSKKKNSDSKETVTIKNSFEASGKENNGSDKKK
 ISNTVEVPKNPKNAVVLVDYGALDVLKELGVADKVKGLPKGENNQSLPKFLDEFKDDKYIN
 TGNLKEVNFQKVASAKPDVIFISGR TANQKNLDEFKKAAPKAKVVYVGTSDNLIKDMKK
 NTE NLGKIYDKEDKAKKINKDLDRKISDMKDKTKDFNKKV MYLLVNEGELSTFGPGGRFG
 15 GLVFDTLGFKPADKKVSKSPHGQINNEYINKQNPDI LAMDRGSVVGGKATTNQVLKNK
 VIKNVKAVKSNHIYELDPKLWYFSSGSSTTTIKQIDELNEVVEKVEK*

Sequence 463
 Contig_0480_pos_3934_4680,
 20 putative peptide of unknown function
 atggataaaataccattaaaagagttagataatttaagtaaaactgataccactgataaaa
 aataaaaaagaatttcgtgcacttcagcaggatattaataactatttgatacctgagttt
 aaaaaatataaaaattccactcaacatttaacagcagataactaatgaggttaagcacttg
 aaagaggattatctaaaaactgttgaaaaataaagagaaatctatatatgatttaaaagaa
 25 tttgtagatttatgtaatcgctcaattaaagataatgaagatattttggattatactaaa
 ttattcgagaaaaatagaactgaagtggagtctgacatttaataaagcacaataaagaa
 gatgcaagtcaacttaaatcaaaattagaagaaaaataatcaacaattaaaagatactgct
 aaaaaatatttaaatcttcaaataatgattctgattcagcgaagaagcaatcaaaaat
 catatttcaccacttattgacaaacaaattacggatattaataaaacaaatatttctgat
 30 aatcatgttgataatgctagaaaaaatgcaattgagatgtattatagtttgcaaaattat
 tatgatacgagagtagatacgattaaaactagcgaaaaattagctcaaattgatgttgaa
 cgattgccccaaaagagggaagatatatcagaaatggataaatcgttcaaaagagaattt
 aaaaaataaaaagaaagtgtaaattaa

35 Sequence 464
 MDKIPLKELDNLSKTDTTDKNKKEFRALQQDINNYLIPEFKKYKNSTQHLTADTNEVKHL
 KEDYLKTVENKEKSIYDLKEFVDLCNRSIKDNEDILDYTKLFEKNRTEVESDINKAQNKE
 DASQLKSKLENNQQLKDTAKKYLNSSNNDSDSAKEAIKNHISPLIDKQITDINKTNISD
 NHVDNARKNAIEMYSLQNYDYTRVDTIKTSEKLAQIDVERLPKEGKDISEMDKSFKREF
 40 KIKESVN*

Sequence 465
 Contig_0480_pos_4829_5149,
 is similar to (with p-value 1.0e-17)
 45 >sp:sp|P39914|YTXJ_BACSU HYPOTHETICAL 12.4 KD PROTEIN IN MUR
 C-ARO A INTERGENIC REGION (ORF2) (ORF3). >pir:pir|S21420|S214
 20 hypothetical protein 2 - Bacillus subtilis >gp:gp|AF00822
 0|AF008220_116 Bacillus subtilis rrnB-dnaB genomic region. N
 ID: g2293135. >gp:gp|X65945|BSAROAG_2 B.subtilis aroA-aroG g
 50 ene. NID: g39812. >gp:gp|Z99119|BSUB0016_49 Bacillus subtili
 s complete genome (section 16 of 21): from 2997771 to 321341
 0. NID: g2635411.
 atggctattaagctgagttcaattgaccagtttgaacaagtattagaagaaaaataatat
 gtttttgattaaaacacagtgaaacttggtccaatttctgcaaatgcgtatgatcaattt
 55 aataagtttttatatgaaagagacatagatggttattatctaatacgttcaacaagagcgt
 aaactatctgattatatacgagagaaaaacacgtaaaacacgaatcaccacaagctttt
 tattttgtagatggtgaaatgaagtggatgcagaccacgatgatattaacgtttctcaa
 cttgctcaagctgaggaataa

MAIKLSSIDQFEQVLEENKYVFLVKHSETCPISANAYDQFNKFLYERDIDGYYLIVQQR
KLSDYIAEKTNVKHESPQAFYFVDGEMKWNADHDDINVSQLAQAE*

Sequence 467.

is similar to (with p-value 2.0e-54)

atgacgattgatagtggaattttatcaaacacctaataacagcttaggggaactatgtatgg

Sequence 468

35

Contig 0480 pos 10463 9366,

is similar to (with p-value $2.0e-25$)

atggttaataaaggagtcggtatggaaatgtttgaagctatcatatataacatatctgtc

120

agtgggtgttgaaaaatcttctttccacctaccaacaaagaagtaatcaagctatcagtt
tcaattggtgtaggatacttaactcaagaagatcgtaaatctcaacgtaaagtattttaa
gatgctgatgacatggtacatgtggctaaaagtgaaggagaataaagtcagtttaaat
cctattgtcaaattataa

5

Sequence 470

MVNBKGVGMEMFEAIYNISVMVAGIYLFHRLQYSENKRMI FSKEYVTVLMTFVSLLLAAY
PIPFQNEYLVHLTFVPLFLGRYTNMIYTLTAAFI VSLVDVFI FGNSIIYGITLIVIAGI
VSAVGPFLLKQNDIISLLILNLISIIILLFLALLSPIYELVEILVLIPISFIITIASAITF
VDIWHFFSLVNRyenEDKYDYLTGLGNVKEFDRHLNEVSSKAEKKQSLALLLIDIDGFK
10 DVNDHYSHQSGDAVLKQMSQLLKNYVPNQFKIFRNGGEFSV VIRDYTLQSVKLAENIR
SGVEKSSFHLPNKEVIKLSV SIGVGYLTQEDRKSQRKVFKDADDMVHVAKSEGRNKVMFN
PIVKL*

15 Sequence 471

Contig_0480_pos_9060_8410,

putative peptide of unknown function

atgaatcgtagtggccatagttatggtttacatgatacatagttttgtgacatcaact
gcaattatcttctcattaaatgatcgtagtagcaggttgattcgtagcgaacgt
20 acaaccgatcttgagaaaaattgctttaaccaatagcctatctcgtaaaatttcgagtaag
caacttacaattgacgaagcaaaaagtgaagtactgcaacttaaacgtgcgtctcttcag
tattcttctttaaacaatctcattgctgcctttgtagcttggtgtttttcttattcatg
tttgggtggcgtagcttccgacgcttggttgcatgcctagcggtggcatagctttttta
acgttttagtttcgtgcaaaaatatatacaaaattaaattcttttcagagtttgtagcatct
25 gctgttggttattagtagtattgcagcaatattcactaaactaggtatagctaaaaatcaagac
attattactattgcaagtgtcatgcctctcggtcccggtattttgattactaacgctatt
cgtgacttacttgccggagagttacttgctggtatgtcacgtggtgttgaaagctgcttta
actgcatttgctatttggtgcaggagtagctattgtattactattattataa

30 Sequence 472

MNRIASHSYGLHDTYSFVTSTAIIFSLNDRSTRLIRIRERTTDLEKIALTNSLSRKISSK
QLTIDEAKSELLQLKRASLQYSFLTNLIAAFVACGFFLFMFGGVASDAWIACLAGGIAFL
TFSFVQKYIQIKFFSEFVASAVVISIAAIFTKLGIKNQDIITIASVMPLVPGILITNAI
RDLLAGELLAGMSRGVEAALTAFAIGAGVAIVLLLL*

35

Sequence 473

Contig_0480_pos_8394_7900,

putative peptide of unknown function

atgtttatcttctggttcactttatcattagtttcattgccacagtccttttttcaatt
40 atatttaaatgcacctaataaattgctattagcttggtgatttggtggagctggtgcttg
acaatatatcagatgacagtaggtatggatttaggtaaagttggcgcttcatttttagga
agtctaataattaggattaatgagtcatacaatgagtagacggtacaagcaacctgttatt
atatttatcgccccggcattatacctctcggtccaggtggcgagcatatgaagctaca
agatttttagtatcaaataattatacgaatgcagttaataacttttttagaggtaacatta
45 attctggtgcaattgcattcggtatacttgtagctgaaatagctattacattttattca
cgcatcaagcaatcttatggtaaaaatcaagggtaaaacttataaaaaatcctataatatg
aataatagagtataa

Sequence 474

50 MFIYLFHFIISFIATVLFSSIIFNAPKKLLACGFGVAVAWTIYQMTVGMDLGKVGASFLG
SLILGLMSHTMSRRYKQPVII FIVPGIIPLVPGGAAYEATRFLVSNNYTNVNTFLEVTL
ISGAIAFGILVSEIVYIIYSRIKQSYGKIKGKTYKKSYNMNNRV*

Sequence 475

55 Contig_0480_pos_7726_6641,

is similar to (with p-value 0.0e+00)

>sp:sp|P55179|PEPT_BACSU PEPTIDASE T (EC 3.4.11.-) (AMINOTRI
PEPTIDASE) (TRIPEPTIDASE). >gp:gp|X99339|BSGALE_6 B.subtilis
orfs 1,2,3,4, pepT and gale genes. NID: gl429253. >gp:gp|Z9

9123|BSUB0020_187 *Bacillus subtilis* complete genome (section 20 of 21): from 3798401 to 4010550. NID: g2636240. >gp:gpID 83026|D83026_30 *Bacillus subtilis* genome sequence covering 1 ic-cel region. NID: g1783231.

5 atggatgaacatgggtacttatttgcctacactcgaaagcaatattaattataatgtacct
actgtcggttttttagcacatgtagacacttcaccagatttcaatgcttctcatgtaaat
ccgcaaatcattgaagcctataatgggcaacctatcaaacttggatgaatctcagcgtatc
ttagatcctgatgttttctgaattaaataaagttgtgggtcatacactaatggtgaca
gatgggtacatctctactaggcgccgatgataaagcaggtgtttagaaataatggaaggg
10 ataaagtatttaattgatcatcctgacattaaacacggtacaattcgagttggctttaca
cccgatgaagaaattggacgaggcccgcatcaatttgatgttagtcgatttaatgcagat
tttgcatatacaatggatggcagtcatttaggagaactacaattcgaaagtttcaatgcg
gcagaggttaactgtcacttgccatggtgttaacggttcacccaggttcagctaaaaatgcc
atgggttaatgcaattagtttaggtcaacagtttaatatgtttacttccctcacatgaagt
15 cctgaaagaactgaaggatacgaaggggttctatcatttaataatgaattttacaggtaatgtt
gaaaaagcaactctacaatatattatcgcgaccatgacaaagaacagtttgagctacgt
aaaaaacgcatgatggaaattcgatgatattaatgttcattataatcattttccaatt
aaagtagatgtgcatgaccaatattttaacatggcagaaaaaattgaacctttgaaacac
atcatgtatatacctaaacgtgtctttgaggcttttagacatcgtaacactgaacct
20 attcgaggtggtacagatggatcaccaattatctttatgggggttacctacacctaatt
tttactggttggtgcaatttccacggtccttttgaatacgttctatcgatgtaattgaa
aaggctgttcatgtgtcgttggtattgtctcaagaagtagcaaacagccatcaatcttat
aaataa

25 Sequence 476

MDEHGYLFATLESNINYNVPTVGFLAHVDTSPDFNASHVNPQII EAYNGQPIKLGESQRI
LDPDVFPKLVVGHITLMVTDGTSLLGADDKAGVVEIMEGIKYLIDHPDIKHGTIRVGFT
PDEEIGRGPHQFDVSRFNAADFAYTMDGSQLGELQFESFNAAEVTVTCHGVNVHPSAKNA
MVNAISLGQQFNSLPSHEVPERTEGYEGFYHLMNFTGNVEKATLQYIIRDHDKEQFELR
30 KKRMEIRDDINVHNFPIKVDVHDQYFNMAEKIEPLKHIIDIPKRVFEALDIVPNTPEP
IRGGTDGSQLSFMGLPTPNIIFTGCGNFHGPFEYASIDVMEKAVHVVVGIAQEVANSHQSY
K*

Sequence 477

35 Contig_0480_pos_3029_2097,
is similar to (with p-value 7.0e-41)
>sp:sp|P18579|MURB_BACSU UDP-N-ACETYLENOLPYRUVVOYLGLUCOSAMINE
REDUCTASE (EC 1.1.1.158) (UDP-N- ACETYLMURAMATE DEHYDROGENA
SE). >pir:pir|S26500|A43727 probable division initiation reg
40 ulatory protein 1 - *Bacillus subtilis* >gp:gp|M31827|BACDDSA_
2 *Bacillus subtilis* (clone lambda-BS1) cell division and spo
rulation protein (dds) gene, complete cds. NID: g142831. >gp
:gp|Z99111|BSUB0008_195 *Bacillus subtilis* complete genome (s
ection 8 of 21): from 1394791 to 1603020. NID: g2633699.
45 atgttcaaaacattgaataaaaatgacatcttacgcgtagtagtcaattcttctctaaa
gatattattaaagtggatgaacctctcaagcgttatatacatatacagaacaggaggagag
gcagattttttatccctaccacaaaatgaagaagtcgaagccatcgtaaagtttgcc
catgagaacagtataccggtaacttatttaggaaatgggtctaacattatcattcgagaa
gggtggaattcgaggaaatcgctctcagcttattatctctcaatcatattgaaacctctgat
50 gatgcaattatagcaggtagtggtgcagcaattattgacgtttcaaatgttgacgtgac
catgtattaaccggttttagaatttgatgcggtatccctgggtcaattgggtggcgccgta
ttcatgaatgctggtgcttatggcggagaagttaaagactgtattgactatgcattatgt
gtcaatgaaaaaggtgatttattaaagctcactacagctgaactggaattagactataga
aatagtggtgtacaacaaaaacatttagttgtattagaggctgctttcaccttagaacca
55 ggtaaattagatgaaattcaggccaaaatggatgatcttactgaaagacgtgaatctaaa
caaccgcttgaaattcccttcttgcggaagtgttttccaaagaccaccgggtcattttgca
ggtaaactcattcaagattctaattacagggctatcgaatcggtggcgttggaagtttca
actaagcatgctgggattcatgttaattgtagacaacggtacagcaactgattatgaagca
cttatacatcacgtacaaaaaattagttaaagaaaattcgatgttggaattgaatactgag

gtacgtattataggtgatcatccacagattaa

Sequence 478

5 MFKTLNKNLILRGLESILPKDIIKVDEPLKRYTYTETGGEADFYLSPTKNEEVQAIKFA
HENSIPVTYLGNLSNIIIREGGIRGIVLSLSLNHIETSDDAI IAGSGAAIIDVSNVARD
HVLGTGLEFACGIPGSGIGGAVFMNAGAYGGEVKDCIDYALCVNEKGDLLKLTAELELDYR
NSVVQQKHLVVLAAFTLEPGKLDEIQAKMDDLTERRESKQPLEFPSCGSVFQRPFGHFA
GKLIQDSNLQGYRIGGVEVSTKHAGFMVNDNGTATDYEALIHVQKIVKEKFDVELNTE
VRIIGDHPTD*

10

Sequence 479

Contig_0481_pos_1175_1876,
is similar to (with p-value 0.0e+00)
>sp:sp|P37478|YYCF_BACSU HYPOTHETICAL 27.2 KD SENSORY TRANSD
15 UCTION PROTEIN IN ROCR-PURA INTERGENIC REGION. >gp:gp|D26185
|BAC180K_1 B. subtilis DNA, 180 kilobase region of replicati
on origin. NID: g467326. >gp:gp|D78193|BACGNTZA_34 Bacillus
subtilis 36kb sequence between gntZ and trnY genes encoding
34 ORFs. NID: g1064780. >gp:gp|Z99124|BSUB0021_146 Bacillus
20 subtilis complete genome (section 21 of 21): from 3999281 to
4214814. NID: g2636442.

atggctagaaaaagttgtttagttgacgatgaaaaaccaattgctgatattttagaattt
aatttaaaaaaagaaggttacgacgtatattgacgttatgacggtaattgacgcagtagat
ttaatctatgaagaagaaccagatattcgtcttacttgatcatgtttacctggtagagat
25 ggtatggaagtgtgctggaagtgcgtaaaaagttatgaaatgccattatcatgctgaca
gcgaagattctgaaattgataaagttttaggtcttgaattaggtgcagatgactacgta
actaaaccatttttagtactcgtgaactcatcgacgtgtgaaagcgaacttacgccgtcat
tattcacaaccagctcaagaagtaagtggtgacgacaaatgaaattacaattaaagatatt
gtgattttatccagatgcatattcaattaaaaaacgtggagaagacattgaattaacgcat
30 cgtgaattcgagctgttccattatcttctaaacatatgggtcaagtcacacgctgaa
cacttactacaaacagtggtgggttacgattatttcggtgatgttcgtactgtggacgta
acaattcgccgtttaagagaaaaaattgaagatgatccatctcatccagaatacattgtg
acacgtagaggcgttggtacttctccaacaacatgattag

35 Sequence 480

MARKVVVVVDEEKPIADILEFNLKKEGYDVYCAVDGNDVLDIYEEEPDIVLLDIMLPGRD
GMEVCREVRKKYEMPIIMLTAKDSEIDKVLGLELGADDYVTKPFSTRELIARVKANLRRH
YSQPAQEVSGATNEITIKDIVIYPDAYSIKKRGEDIELTHREFELFHYLSKHMGMQVMTRE
40 HLLQTVWGYDYFGDVRTVDVTIRRLREKIEDDPSPHEIYVTRRGVGYFLQQHD*

Sequence 481

Contig_0481_pos_1889_3721,
is similar to (with p-value 0.0e+00)
>gp:gp|D78193|BACGNTZA_33 Bacillus subtilis 36kb sequence be
45 tween gntZ and trnY genes encoding 34 ORFs. NID: g1064780. >
gp:gp|Z99124|BSUB0021_145 Bacillus subtilis complete genome
(section 21 of 21): from 3999281 to 4214814. NID: g2636442.
atgaagtggcttaaaacaactacaatcccttcacacgaaactcgttattgtttatgtacta
ctcattattattggtatgcaaatcatcggtttgtattttacgaatagtttagaaaaggaa
50 ttactcgataacttcaagaagaacataacacaatatgcgaagcaattagacgtcaatatt
gaaaagggtttataaagataaagataaagggttcagtcacacgctcaaaaggatatccaagac
cttttgaatgaatatgcgaatcgcaagaaataggagaaatagcgtttattgataaagac
caaattatcatggcaacaaccaagcagtcctaaccgtggtcttatcaatcaaaagggttaac
gacggttcagttcaaaaggcgctctccttagggcacaacgaatgatcatatggttcttaag
55 gattacggaagtggtaagagcgtgtttgggtatataatataccggttaaaagttgataaa
cagacaatcggtgatataacatagaatcgaaaattaatgatgtatacaatcagctgaac
aacattaatcagatattcatcgtaggacagcgatattcattacagtaataacta
ggattcttcatgcacgaacgattactaagccgataaccgatatgcgtaaccaaacggtt
gagatgtctaaaagtaactacacgcaacgagtgaaatatacggtaacgatgaaatcgggt

gagctcgcaacttgccctcaataaacttatcgaaacgtgtccaagaagcacaagcgaataca
 gaaagtgagaaacgtcgcttagattctgttatcacacatatgagcgatggattcttgccg
 acagatcgccgtggacgtgtacgtattgcaaacgacatggcgctgaaaatgctcggtctc
 gcgaaagaagatgtcatcggtactacatgcttgggtgtccttaacttagaaaaatgaattc
 5 tcattagaagaattcaagaaaatagtgttcttctgttagatattaacgaagaagaa
 ggcatattgacacgtgtaaacttttagtacgattgtacaagaacagggtttcgtgacagg
 tacattgccgtactacatgatgtcacagaacaacaagtcgaacgtgaacgtcgcgaa
 ttctgttgcgaatgtttcacatgaattacgtacaccactgacatcgatgaatagctatctc
 gaagcacttgaagaagggtgcttggcaagataaagaactggcaccatcattcctatctgtc
 10 acacgcgaagagactgaacgtatgattcgttttagtgaatgatttacttcaattatctaaa
 atggataatgaatcagatcaaattacgaaagaattatcgacttcaatatgtttatcaac
 aaaattattaaccgtcatgaaatgacagcgaaagatacgacattcgtacgcgaaattccg
 caacaaaccattcttgcgtgaaatcgatccagacaagatgacacaagtatttgataatgtc
 attaccaatgcaatgaaatattcacgtggcgagaaacgtgttgagtttcatgtgaaacaa
 15 aatgcactttacaatagaatgacgattcgtattaaagataatgggtattggaataccgatt
 aacaaggtagataaaaatatttgatagattctatcgtgtagataaagcacgtacacgtaag
 atgggtggtacaggactaggcttagctatttccaaagagattgtcgaagcacataacggt
 cgaatttgggctaacagtggtggaaggacaaggtagctcaatctttattacacttccttgc
 gaaatcattgaagacgggtgattgggatgaataa

20 Sequence 482
 MKWLKQLQSLHTKLIVIVYVLLIIIGMQIIIGLYFTNSLEKELLDNFKNITQYAKQLDVNI
 EKVKDKDKGSVNAQKDIQDLLNEYANRQEIGEIRFIDKQDIIMATTKQSNRGLINQKVN
 DGSVQKALS LGQTNHMLKDYGSGKERVWVYNIPVKVDKQITIGDIYIESKINDVYNQLN
 25 NINQIFIVGTAISLFITVILGFFIARTITKPITDMRNQTVEMSKGNYTQVRVKIYGNDEIG
 ELALAFNNLSKRQVEAQANTESEKRRLLDSVITHMSDGLATDRGRVRIANDMALKMLGL
 AKEDVIGYYMLGVNLNEEFSLEEQENSDFSLLDINEEGIIARVNFSTIVQETGFVTG
 YIAVLHDVTEQQQVERERREFVANVSHELRTPLTSMNSYIEALEEGAWQDKELAPSFSLV
 TREETERMIRLVNDLLQLSKMDNESDQITKEIIDFNMFINKIINRHEMTAKDTTFVREIP
 30 QQTIFAEIDPDKMTQVFDNVITNAMKYSRGEKRVEFHVKNALYNRMTIRIKDNGIGIPI
 NKVDKIFDRFYRVDKARTKMGGTGLGLAISKEIVEAHNGRIWANSVEGQTSIFITLPC
 EIIEDGDWDE*

Sequence 483
 35 Contig_0481_pos_3765_5051,
 putative peptide of unknown function
 atgagtatcggttttgacatacatgggtctggaacttttctccagacctttcaaatattgat
 aacacggataatagtaaaagtgtataagcctaaaccacttactaaaccaatgactgcagaa
 atggaaggaacgattacaccatttcaaatcggtgcattctagagatgaaaaatctcaagga
 40 acagtggcatcaggtgcagtccttagacaagatgattcaacctttaaataatcaagaagtt
 aaatctgtatcacatctgaaaagggaacataaccttgttatacctgaactaagcaacgac
 ttatctgtccttagatttcaacttatgatttgccactttcaacatacttaagtcaagtactc
 gatatcgatgcgaaagtgccgaataactttaattttgatcgccctccttatcgatcaagat
 cataataaccacgtcgactatattgcgatttagcaaagaccgtcatgaagtagttaactt
 45 aagacaacgatgaaagggaataacggtgacaaagcttttaaagtagtcgaacctgacatg
 caaccctatacggaaatcatcacgaataaagatacaatcgacaaagcaacacacgtgtt
 gcaccaagcaaacggaaagacttaagacgtatcgcatgggtcttcaatacgtacgtgtt
 gaacgcatagaactcaatactatttgatgattcaacgattgttcgtagctctcaaagtgg
 acgacaacatacaacaataactgggtgcgccaactataacgataaagatgaaatgtat
 50 cattataagaatttatctgaagacgcgaaaagtcaagcaacatgcaagaaaccatccca
 ggacatacagatttataaatagtcattgggtggtctttaaataagattatcgccatttt
 aagacagataatagaacggggaaactcacatatcaaagattcctcaacggtcacccaacg
 ttaataaacataacttcaatgaaatccaagtcacatggggggataaaggcggttacgat
 tatcaacggttcgctacttaagacggacgtcacactgaacagtgaagaatctaaatccgtc
 55 ctaaccggttgagtcggtggttctgcattagccaacctcctgatattgattttgaaaag
 gtaacgaacattgcgattgggttatgatggacgacaaggcaataacgaagatattgaa
 gttcaacgtaactgtgaattaataaccacgttggtttgtagaatacgtggcaattgggtat
 gcctataaagatggggaggcttgaataa

Sequence 484

MSIVLTYMVWNFSPDLSNIDNTDNSKSDPKPLTKPMTAEMEGTITPFQIVHSRDEKSQG
TVASGAVLDKMIQPLKNQEVKSVSHLKRHNHVIPELSNDFIVLDFTYDLPSTYLSQVL
DIDAKVPNNFNDRLLIDQDHNHVVLYAISKDRHEVVKLKTTMKGNNVDKAFKSI EPDM
QPYTEIITNKDIDKATHVFAPSKPKDLKTYRMVFNTISVERMNSILFDDSTIVRSSQSG
TTTYNNNTGVANYNDKDEMYHYKNLSEDAKSSSNMQETIPGTYEFINSHGGFLNEDYRLF
KTDNRTGKLTYYQFLNGHPTFNKHNFEIQVTWGDGKVYDYQRSLLKTDVTLNSEESKSV
PTVESVRSALANHPDIDFEKVTNIAIGYDMDDKANNEDIEVQRNCELI PRWFVEYDGNWY
AYKDGRLE*

Sequence 485

Contig_0481_pos_5052_5843,

putative peptide of unknown function

atgaactggaaactcacgaaaacacttttccattttcgtttttattcttctgtgaacatcttt
ttagtcatcgtttatattgataaagtgaataaatcacaagttaatgactcggaaaaggta
aacgaggtcaattttcaacaagaagaattgacgtgcccaaggatgtcttgaatcaaaat
gttaaagatactgaacttgaacaaattactgcccggttcaaagaatttctcaagttatgcg
aaagtcattcattcaagcatgcaaacgtctgattccgacaaaacacttgaaggagatattgat
aaaggcgttcaagtgagtgataaagaacttacaagatatcaaagagtagcattgcaaagaaa
atctttaacggttaaagagtatcagtttaagtgttttaactaaagataaagtcacttacgaa
caaacgtataaagattatccgattatgaataatagtaaagcgcgctaacgtttaatttg
agcgatggcaaggcgacaagctataaacagacagcgatggatgatatacaagtagctaaa
ggttcaaatagcacgaagaacaagtcacacgcccgttaaagctattgaagccctttat
tacaatagatatttaaaacaaaatgatcaagttcttgatgcacgcctaggtattattca
gttgtaaaggaaacaaacgttcaattactccaacctaactgggaaattaaagtaaaacat
aaaggcaaggatgaagttcaaacctattatgtagaagctacaaatcataatccgaaagtg
attgattattag

Sequence 486

MNWKLTKTLFIFVILVNI FLVIVYIDKVNKSQVNDSEKVNNEVNFQEEIDVPKDVNLQN
VKDTELEQITARSKNFSSYAKDHSSMQTSDSDKTLEGDIDKGQVQSDKNLQDIKEYIAKK
IFNGKEYQLSDLTKOKVTYEQTYKDYPIMNNSKARLTENLSDGKATSYKQTAMDDIQVAK
GSNSTKKQVITPRKAIEALYYNRYLKQNDQVLDARLGYYSVVKETNVQLQPNWEIKVKH
KGKDEVQTYVEATNHNPKVIDY*

Sequence 487

Contig_0481_pos_6579_7265,

is similar to (with p-value 8.0e-79)

>gp:gp|D78193|BACGNTZA_30 Bacillus subtilis 36kb sequence be

tween gntZ and trnY genes encoding 34 ORFs. NID: g1064780. >

gp:gp|Z99124|BSUB0021_142 Bacillus subtilis complete genome

(section 21 of 21): from 3999281 to 4214814. NID: g2636442.

atggaagaacttttcagccaaatcgacagaaacattaaggatttaaacggaatttttagtg
acacatgaacacatcgaccatattaaagggtcttggtgttttagcacgtaaatataaactt
ccgattttacgcgaatgagaatacatggaaagcgatagagaagaaagatagccgcattcca
atggatcagaaatttatctttaatccatatgaaacgaaatctcttcaggatttgatata
gaatcatttaacgtgtcacatgacgcgattgatccacaattctacatcttcacacaataac
tataagaaatttacgatgataactgacactggttacgtttcagatcgatgaaaggtatg
attcaaggtagtgatgtctttatgtttgaaagtaatcacgatgtcgatatgttacgcatg
tgctcgctatccatggaagacgaaacaacgtatttttaagtgatatgggtcacgtatccaat
gaagacgcgggtcttgcatgagtgatgtcattacaggttaatacgaacgtatataacctc
tctcattttgtcacaagacaataatgaaagacctcgacgcgatgagtggtggacaagtg
ctcaacgaacacgatatcgatacagagaagaagattgtctttgcgataccgataaagca
caagccacaccgatttatacactataa

Sequence 488

MEELFSQIDRNIKDLNGILVTHEHIDHIKGLGVLARKYKLPIYANENTWKAIEKKDSRI P
MDQKFIFNPYETKSLAGFDIESFNVSHDAIDPQFYIFHNHYKKFTMITDTGYVSDRMKGM
IQGSDVFMFESNHDVMDLRCRYPWKTKQRILSDMGHVSNEAGLAMS DVITGNTKRIYL

SHLSQDNNMKDLARMSVGQVLNEHDIDTEKEVLLCDTDKAQATPIYTL*

Sequence 489

Contig_0481_pos_8820_7690,

putative peptide of unknown function

5 .atgaattcgttgacatagcaggacgtattttcaaacagacgattcgagatgtaagaaca
ttggcactgttacttattgcacctatattactattgtcgctactatattacatttttaca
gttgccgataatacgaatggcgtaacagttggggttcacgatgtaccagattcattaatg
actgaattacatgataaagatattcacgttaaacattataaaaatgacaatgatataagt
10 gataaaattaaagacgacaaaattaacaggatTTTTGcacagtgatggTcaaaaagtatca
gtgacttatgctaacgataatcctacacaagcaggagaaactaacaggTgcaaatcaaaaa
tggttaatgagtcataacatgaatgcatgaaagataactaataaattgcatcaagcg
ttaactaaaaatacaacaaaaaatgcccggggatgggggagacacgcctcatcaagatatg
gctaaaccatataaaactaacaacgcactatttatatggTtcatcagattctacgtatttt
15 gatatgataaaatcctatttttaattggattttttgtctttttctttacgttttttaatttct
ggcattggcttatttaaaagagcgctacttctggcacattagaacgtttacttgccctctcca
ataaaaagaagtgaattatttttgggttatgttttcggttatggtagtttttagcgttatc
caaacaatagttgtcgatttatatgcaatttatattctgcatatagacttagtaggttcg
atatggttcgtactattaacggcaatattaacagcgctgtcgctgtgacattcggtata
20 ttattatctaccttttgcttcctcagaattccaaatgattcaatttataccattagtcata
gtgccacaagtactatttgaggcattataccaattgaatcaatgaataaaggattacaa
tacttttcacatatcatgccgttattctataccggccaaacgatgcaaaatattatgatc
aagggttatggattcaacgatatttacatttatttaattgtgttattcgcatttttcatt
ttcttattgattttaaatattatagggcatgaaaagatatagaaaagtttag
25

Sequence 490

MNSLHIAGRIFKQTI RDVRTLALLLIAPILLLSLLYYIFTVADNTNGVTVGVDVPSLM
TELHDKDIHVKHYNKNDNDISDKIKDDKLTGFLHSDGQKVSVTYANDNPTQAGELTGANQK
30 WLM SHNMNAMKDN TNKLHQALTKIQKMPGDGGDTPHQDMAKPYKLTHYLYGSSDSTYF
DMINPILIGFFVFFFTFLISGIGLLKERTSGTLERLLASPIKRSEIIFGYVFGYGSFSVI
QTIVVVLYAIYILHIDLVGSIWFVLLTAILTALVAVTFGILLSTFASSEFQMIQFIPLVI
VPQVLFAGIPIESMNKGLQYFSHIMPLFYTGQTMQNIMIKGYGFNDIYIYLIVLFAFFI
FLILNIIGMKRYRKV*

Sequence 491

Contig_0481_pos_7680_7336,

putative peptide of unknown function

35 atgaaccaagatattaagtcatttagttgaaaccattgtgcctcaacttgaatatttaagc
gataaacaagacgtgtcatagaaagtgtcattgtcattattcagtgaacaaggatttgat
40 aaaacgagtactaaagaaattgcgagcggtgcaaatgtcgagaaaggaaacgggtatttaag
cagtttaaaagtaaaagaatgttattatatataaatcacaaagcgtgtaagacatcggtc
tcccccttccatgatgcgctttcatttcaaaaaattccttaataatcggttcagtcgca
agcctgttttaactaacacataatcattccctgcaattctcctga

Sequence 492

MNQDIKSLVETIVPQLEYLSDKQRRVIESAIALFSEQGFDKTSTKEIAQRANVAEGTVFK
QFKSKRMLLYINHKACKTSFSPFHDALSFQKNSLINRSCASLFLTHNHSLOFS*

Sequence 493

Contig_0482_pos_2955_2551,

putative peptide of unknown function

50 gtgtcagtaactgtataaaggacaaaactgaaacagaatggcttccagttattggattttaga
aaciaatccttagcaaaagggtagcgcgacaacatttgatattaataaagctcaaaaacgt
55 tgtttcgttaaaagctgcagcattacatggcctaggtctttatatatacaacggggaagaa
gttccaagcgctaacgacaatgacattacagaattagaagagcgatcaaccagtttgta
acttcatctcaagaaaaaggttagagacgcaacgctagacaaaacaatgcgttggttaggt
attcaaaacattaaacaaagttactaaaaaagatatagcaaatgcacatcaaaaactagat

gcaggactaaaacaattagataaggagaattcaaatgttaaatag

Sequence 494

VSVTVKGQTETEWLPVLDFRNKSLAKGSATTFDINKAQKRCFVKAAALHGLGLYIYNGEE
5 VPSANDNDITELEERINQFVTSSQEKGRDATLDKTMRWLGIQININKVTKKDIANAHQKLD
AGLKQLDKENSINVK*

Sequence 495

Contig_0482_pos_2104_1430,
10 putative peptide of unknown function
atggttagtaataaaaaactacattacagaagatgacggtacaacaactgtagtcatcaaa
ggagtagaactagataacaaaacatctttacttttagacaacggttacgaagtagaagca
gatgtaagagttgtagatccattcaagattacagataagcagcgtagaaaagtatttgct
ctctgtaacgacatagaagcttacacaggacaaccacgcgactatatgaggtatttgctc
15 atggattacgtagaagttctctatggctatgaaaaacgtctctcattgagtgattgcaca
agagaacaagctaaacaagttatagaagttattcttgactgggtgtttcacacaatatata
ccacttaattataagacaagtgaacttactcaaaaatgataaagcggttcctttactggtca
acagtcaatcgtaactgtgtaatatgcggaacgccacgacgagaacttgcgcatatcac
acagtaggctcgaggacgtaacagacgaaagatagatcacacagacaacaaagtattagcg
20 ctatgttcaagacatcataaagagcagcaccaaataggtatagatagttttaatgagaaa
tacaattacatgaaagttgggtgtccgtagatgaacgactcaaccgaatgttgaaagga
gaagtaaatggctga

Sequence 496

25 MVVIKNYITEDDGTTVVIKGVELDNKTSLLLDNGYEVEADVVRVDPFKITDKQRRKVFA
LCNDIEAYTGQPRDYMRYLFMDYVEVLYGYEKRLSLSDCTREQAKQVIEVILDWVFHNNI
PLNYKTSDLLKNDKAFLYWSTVNRNCVICGTPRAELAHYHTVGRGRNRRKIDHTDNKVLA
LCSRHHKEQHQIGIDSFNEKYKLHESWVSVDRLNRMLKGEVNG*

30 Sequence 497

Contig_0482_pos_1401_643,
putative peptide of unknown function
atgttcgatgatagcaaaatcaagtatatagaagcactgccagaacgagatacaatcatc
actttatgggttaagtgtgctgacattagctggaaagtataacgaacaaggatacattatg
35 ttatccgaaagtctaccctataacgaagaaatgttagctaacgaatttaatagacctatc
aattcaataagattagcgttacaaacattcgaaaagctaagcatgattgaagaagtgaat
ggtgtctttaaagtatctaattgggaaaaacatcagaacatcgaaggtttagaaaagata
agagaacaaaaccggtttgcgtaaacaaaagcaaaagaaaaaacaaaacttttagatagt
cacgtgaagtcacgtgacagtcacgcaacagatatagaagaagataaagaagtagaagaa
40 gaaaagagaaaaagaagtagataaagatatcttcaaaaactcaattaattacatcatgagt
aaccttactcataatttaactcctaaccaaatggaacagataggatatgccattgatgat
attggacaacatgcagatgaagttgttgaagtagctactgattatacaaaagacaaaggt
tgtcatgcaggttacctaatacaagtggttaaacactgggctaaagagaacgttaagaat
aaaaaagaggctgaaaataaaattaaacctaaaaataaaaaaactgtaacagatgatgta
45 attgctcaaatggagaaagagctaggagatgaaagttaa

Sequence 498

50 MFDDSKIYIEALPERDTIITLWVKLLTLAGKYNEQGYIMLSLSESLPYNEEMLANEFNRPI
NSIRLALQTFEKLMSIEEVNGVFKVSNWEKHQNIIEGLEKIREQNRLRKQKQKQKLLDS
HVKSRSRSHATDIEEDKEVEEEREKEVDKIDFKNSINYIMSNLTHNLTPNQMEQIGYAIDD
IGQHADEVVEVATDYTKDKGCHAGYLIKVLNNWAKENVKNKKEAENKIKPKNKKTVTDDV
IAQMEKELGDES*

55

Sequence 499

Contig_0482_pos_637_284,
putative peptide of unknown function
atgactaaacaacaagccctagaagtaattaagacaattagacatgtatacaacattgac

tttgacagacctaattagaaacatgggttaacattttgagccaaaatggggattatgaa
 ccgactaaaaaacagtaatgcaatatatcaatgatgctaataccttatccacctagtatt
 ccaaacataatgagaaaagaagtcagagtcgtaaaagaagagcctgtcgacgaaaaaact
 gctagacatcggttgagaatgaaaaatgatccagaatacgtagcacaacgtaaaaagata
 5 ttagacgacttcagaaaagaagttaagtgaagtttgagtgagtgacgatgaatga

Sequence 500

MTKQQALEVIKIRHVYNIDFDRPKLETWVNILSQNGDYEPKKTVMQYINDANPYPPSI
 PNIMRKEVKVVKKEPVDEKTARHRWRMKNDPEYVAQRKKILDDFRKKLSEFGVSDDE*

Sequence 501

Contig_0483_pos_6911_7564,

is similar to (with p-value 1.0e-41)

>gp:gp|U93874|BSU93874_2 Bacillus subtilis cysteine synthase
 15 (yrhA), cystathionine gamma-lyase (yrhB), YrhC (yrhC), YrhD
 (yrhD), formate dehydrogenase chain A (yrhE), YrhF (yrhF),
 formate dehydrogenase (yrhG), YrhH (yrhH), regulatory protei
 n (yrhI), cytochrome P450 102 (yrhJ), YrhK (yrhK), hypotheti
 cal protein YrhL (yrhL), putative anti-SigV factor (yrhM), R
 20 NA polymerase sigma factor SigV (sigV) and YrhO (yrhO) genes
 , complete cds, and YrhP (yrhP) gene, partial cds. NID: g193
 4604. >gp:gp|Z99117|BSUB0014_205 Bacillus subtilis complete
 genome (section 14 of 21): from 2599451 to 2812870. NID: g26
 34966.

25 atgacacctttgggtcaatcaccttttagcactagcgctgacatagttattcatagtga
 actaaatttctaggttgacatagcgatttaattgcaggtgcagcaattactaataataga
 gaggttgcaaatgcattgtactttattacagaacggcacgggcacagccctttctgcatat
 gatagttgggcacttgcaaaacatcttaaaacattaccagttcgttttaacaatctgtt
 cataatgctgaacgccttgttcaatttttgagtcagagagaggagatttctgaggtgat
 30 taccgggaaataatcttacacatctcaagcaagcttcaactggaggtgcagtgataggt
 ttccgacttaagatgaatctaaagcacaaaagttcgctgattctcttactttaccactt
 gtatcagtgagtcctcggtggtgtagaaactatcctatcacatcccgcaacaatgtctcat
 gcagcagtgccagaagatgtgagacgtgaacgtggcatcactttcggttattccggtta
 agtgtaggtcttgagaattcagaagaactcatcgagattttaactacgctttaagggag
 35 gctttcaatgagtcatttactgaaccaattaaagagcaacgtttttagtagctga

Sequence 502

MTPLGQSPLALGADIVIHSATKFLGGHSDLIAGAAITNNREVANALYLLQNGTGTALSAY
 DSWALAKHLKTLPVRFKQSVHNAERLVQFLSQREEISEVYYPGNNLTHLKQASTGGAVIG
 40 FRLKDESKAQKFVDSLTLPLVSVSLGGVETILSHPATMSHAAPEDVRRERGITFGLFRL
 SVGLENSEELIADFNALKEAFNESFTEPIKEQRFSS*

Sequence 503

Contig_0483_pos_14208_0,

is similar to (with p-value 0.0e+00)

>gp:gp|AB015981|AB015981_5 Staphylococcus aureus genes for O
 45 rfa, MnhA, MnhB, MnhC, MnhD, MnhE, MnhF and MnhG, complete c
 ds. NID: g4001723.

atgtttatgttgattggtattattgggtcattttacaacaggagatattttcaacttgttt
 50 gtgttctttgaagtcttttaagtcttcatattgtttactcgttattgggtactactaaa
 atacaattacaagaaacaattaagtatatttttagtcaatgtttgtttcatcgtctttcttt
 gtcattgggtgttgagttttatattcagttgttaggaactttaaatctcgctcatattagt
 gaaagattgtcacaactttctgtacatgacagtggttagtcaatattgtttttatttta
 tttatctttgtcttttgccactaaagcagcgctttttcctatgtacgtatggctacctggg
 55 gcttattatgccccctccagtagcgatcatcagttctttgggtgcactattgactaaagtg
 ggtgtatacgcaattgcgagaactctaagtttattcttttaataatacagtaagcttttct
 cattatgtcatccttttcttagcattacttacaattatttttggtatgtaggtgcgata
 gcttactatgatacgaagaaaatcatcctttacaatattatgattgcagtaggtgtcata
 ttagtgtgtattgctatgatgaacgaatcaggcatgactggtgcaatatattacacacta

catgatatgtagttaagcttcattgttcttactcattggcgatcatgtacaaaatcact
 aaaacgactgacttacgtcattttggtggcttgataaaaggtatcctattctaggttg
 acattctttattgcagcgctaagcttagcgggtataccaccttttagtggtttctacggt
 aaattctatattgttcgagcgacctttgaaaaaggattttatctaagtggtatcattgta
 5 cttttatcaagtttaactcggttatattcagtcatacgtattttcttaaaaggatttttc
 ggtgaagttgaaggatatactttatctaaaaaggtaaatggttaaatatctaacaactatc
 gctgttgcatctacagt

Sequence 504

10 MFMLIGIIGSFTTGDIFNLFVFFVFLMSSYCLLVIGTTKIQLQETIKYILVNVVSSSFF
 VMGVAVLYSVVGTNLNAHISERLSQLSVHDSGLVNIVFILFIFVFATKAGVFPYVWLP
 AYYAPPVAIITFFGALLTKVGVIARTLSLFFNNTVSFSHYVILFLALLTIIFGCIGAI
 AYYDTKKIILYNIMIAVGIVLGIAMMNEGSMGTGAIYYTLHDMLVKASLFLIGVMYKIT
 KTTDLRHFGGLIKGYPILGWTFIAALSAGIPPSGFYKGYIVRATFEKGFYLSGIIV
 15 LLSSLIVLYSVIRIFLKGFFGEVEGYTLSKKVNVKYLTTIAVASTV

Sequence 505

Contig_0483_pos_13585_12029,

putative peptide of unknown function

20 atgaacgatttataccagcttggtatagccgtaacagatgggtgggaaagtacctcaaga
 ccattctatctaaaaaacagtatacagattttgacgatatgattagtttaatgacaatg
 catagttcgaataatgtggattatcaattgatagttttaaatttttagtccatatcttaga
 acattcctccatcgatatgatttgtatgaaagtcattattggctctgattttgatgagata
 cagggcggttgacatcaaacgcctcaagctattgattatcgcgatctttcatggccagaa
 25 ggcactgaatttttttactccctttcaaattcaagcgattacaggtgataaacagttt
 tctaaaaattcacttcagccaagaggggtacctgatgtgggtagaggattacaagtatagt
 acaattcaaaagcagattttgtattcgatgacagaggatttatatcggcagtgcgacttat
 acacctgatggtgataacaataaaaaacactatttttcaaaagatggggaagaaatattt
 gttgaagacttaaatgttaatacagtaacgattaataaaaaattccaatcaaaatttaaa
 30 agggttacgtattcatctatggctgagttgatagaagagaaattccaatcatatgtagaa
 agagaattgaatgaagatgattctgttatagtgccatctgatgaacgtcataattcaatg
 atggcacgcactattgatgcatcgctttatgttttctatttttactgagagaaataaa
 gtggtgacacaagaatttatatgactctatttctagagcatattattgtctcggtgatata
 caagctaatacaaaatgatgtgaacactacgcaggattgaacatgaatgatattaatctt
 35 ttaagggttaacgccttttgatgcgaagtcattacctaaccaagtagtcaattgtatgac
 acttatattggattatggatagatggtttgacgagattgaaatacgagagattgtaaac
 agcttatttcaatatattcaacataaagatggctataagttgaaaattttaactaagagt
 agagataatcttacggaaaatcttatagatgaagttgctcatctcaatgatttatatcac
 caagagaaaaaggaataaagtgatgtaattgaagacgtgatacagaataaaaaagaaaca
 40 atcattgatattgaacagtagcgtttgagaagatcttgtaagcgttatttcaaaatta
 agagttgtagtagatttatctttagagccgaaacttttttacaatatctgttgattggc
 gcgggtataccacaaattaataaaaaagagaacagattatgttaaacatatgcataatgga
 tatattattgatgacatatcgcaaactgtagaatctttagattattttttggcacattta
 aaaaattggaattattcttatgcatattccatgagattaacggatgatttttagttcaatt
 45 aatattattcatcaaatatcagttatttaaaagtgatgtttcaagtggcacgtaa

Sequence 506

MKRFIPAWYSRNRWESTSRPFYLLKQYTFDDMISLMTMHSSNNVDYQLIVLNFSPYLR
 TFLHRYDLYESHYWSVFDEIQGVGHQTPQAIDYRDLSPGTEFIFTFQIQAITGDNTF
 50 SKIHFSQEGYLMWVEDYKYSTIQRFRVFDGRFISAVRTYTPDGDNNKKHYFSKDGEEIF
 VEDLNVNTVTINKNFQSKFRVTYSSMAELIEEFQSYVERELNEDDSVIVASDERHNSM
 MARTIDASSLCFSIFTERNKVVTDQDLYDSISRAYCLVDTQANQNMIEHYAGLNMNDINL
 LRVTPFDAKSLPNQSSQLYDITYIGLWIDGLDEIEIREIVNSLFQYIQHKDGYKLKILTKS
 RDNLTENLIDEVAHLNDLYHQEKKEISDVIEDVIQNKKETIIDIETVPFEEDLVSVISKL
 55 RVVVDLSLEPKLFLQICIGAGIPQINKKRTDYVKHMNGYIIDDISQTVESLDYFLAHL
 KNWNYSYAYSMRLTDDFSSINIIHQINQLFKGDVSSGT*

Sequence 507

Contig_0483_pos_11796_10480,

- putative peptide of unknown function
 gtgattgacaatgagtgattgggataatcaataccaacaagataagacaataacaacgtaat
 ttataaaaaccactcatttatgaaaatgaagaacaattacaacaaaaactagaggcagtt
 acatttcctgggcaatatggagataaagttaaacctattcattgtcgcgtagtattcat
 5 ttgatgggttcttatcaatttaattgaaaatgagtgctattgaagtatcaggacgatttggg
 gaatcataccaacccctcattacatggagtcacaaatatcattgctgatgccaataagggtg
 aatcaaatatggccagaatttaaaagttgaaggtgatgctaaaatccaatatacattgaga
 ttgacgcctgtttattcaactgatcaaccagtagaaaagctaataatatgaacaagacgat
 ttagacactcccatagaactacctgctcgtccttatcaaacatatgtgagtgatcaatc
 10 aaagctaaaggtaaagggaacattatattataggtgctattcataaacgttgggtcacgcttg
 gaattagggcagttcatatttaggcggaaaacgatatagtgatgaaaataagcaagaattt
 atacattactctcaactcctggagattttaaaccaccactcaatgtatattttagtgggtat
 cgtactgctgagggtcttgaagggtactttatgatgaaacgatgaatgctccatttatt
 ttaatagctgatcctagaatcgaaggtgggtgccttttacctagggtcagagaattatgaa
 15 caggcaatccgtaaggtcatccaaatgctttggattatttgggatttgcgaacaaccaa
 ttaattctttctggattatcaatgggatcatttggcgcactttattacgctacaaaatta
 aatccagcggctgttattgtaggaacacctttgataaatctcgggtactattgctaataat
 atgaaactcgttcgtccaaacgattttggaacgtcacttgatattttgcgattgaatcaa
 aatggcataactcaacaaagatgttggtagttagataatcatttttgggaagcaaatcag
 20 catagtatttgtcaatgaccacatttgcgattgcttacatggagcatgatgattatgac
 aaatatgcatttcaagatttattgcctgttcttataaaaacaacatgcacgtgtgataagt
 aaaagaattcctggtagacataatgatgattctgctactgttactcattgggtttattaat
 ttataaatttaatcatggaagagcgatttgggagggtaacacatgcaagaagatag
- 25 Sequence 508
 VIDNEYWDNQYQDDKTIQRNFIKPLIYENEEQLQQKLEAVTFPGQYGDVKVPIHCRVSIH
 FDGSYQFNGNESIEVSGRFGESYQPLITWSQNIADANKVNQIWPEFKVEGDAKIQYTLR
 LTPVYSTDPVEKLIYEQDDLTPIELPARPYQTYVSVSIKAKGKGTLEFIGAIHKRWSRL
 ELGQFILGGKRYSDENKQEFIHYPGDLKPPLNVYFSGYRTAEGFEGYFMMKRMNAPFI
 30 LIADPRIEGGAFYLGSENYEQAIRKVIQNALDYLGFANNQLILSGLSMGSFGALYYATKL
 NPAIVGVKPLINLGTIANNMKLVPRNDFGTSLDILRLNQNGITNKDQVQLDNHFWKQIQ
 HSDLSTTFATAYMEHDDYDKYAFQDLPVLTQKHARVISKRI PGRHNDDSATVTHWFIN
 FYNLIMEERFGRVTHARR*
- 35 Sequence 509
 Contig_0483_pos_10436_9921,
 putative peptide of unknown function
 atgtatggtacaaaattacgttttaatacaagataatatctattttgagaaccctttgatg
 ccattccggtacaaatcattcacagttggatatgttaactgattttgcagaagaccgtgta
 40 agccctaagctacctatttttaaaaaaaggcgccaatatcaatttcaatttaattttgaa
 gttgaacctgaggggtgcggcttattttaaaatgaaattttatcgtaagaataaagaaatt
 cttagtcatacaattctaaaaataaaaaaagaaaatattgtctatcctagagaagcatat
 tcatatgaattagaacttattaatgctggcatgaatcatctatctttcacaaatataatt
 gtgcaagaattaagagaagatagtaatacaagcttatgaggcaacgcaatatatagatcct
 45 aagaaaaaacttaagtaattaatacaataataaccaatataaggacacatcatctagac
 tcatcaactatcacaggagtgatgatgaatggctaa
- Sequence 510
 MYGTLRFNQDNIYFENPLMPSTIIHSWYMLTDFAEDRVSPKLPILKKGRQYQFQFNFE
 50 VEPEGAAYFKMKFYRKNKEILSHQILKNKKENIVYPREAYSYELELINAGMNHLSFHNII
 VQELREDSNQAYEATQYIDPKKKLKVINQIITNIRTHHLDSSNYHRSDMNG*
- Sequence 511
 Contig_0483_pos_9697_9095,
 55 is similar to (with p-value 4.0e-42)
 >gp:gp|X62035|BSSECA2_1 B.subtilis secA gene (partial). NID:
 g48979.
 atgtatccaaaagatgtgcagatttttaggagcaatcgctatgcatcaggggaatattgca
 gaaatgcaaacaggagaaggtaagacgcttacagctaccatgcctctgtacttaaatgca

cttacaggtaaaggtgcttatctaatacacaacaaatgattacttagcaaaacgcgatttt
 ttagaataaaccactatatgaatggctaggcttgtctgtatcattaggatttgtggac
 attccagaatatgaatacgtgaaaatgaaaaatatgaactgtaccaccatgacattggt
 tacacgactaatggcgactagggtttgattatttaattgataatttagctgatgatatt
 5 cgtgccaaatttttaccgaaattaaactttgctattattgatgaagtcgattctattata
 ttagacgctgcccaaacgccttttagttatttctggtgcaccacgtgtacaatctaattta
 tttcataaacttaattcttttagtcttttctttatcatctttaacagctttgaattggtta
 agtacatcgtctgacattttgatgccaggcacttcattatgtaagaaaagtgcgttgtaa
 taa

10

Sequence 512

MYPKDVQILGAIAMHQNIAEMQTGEGKTLTATMPLYLNALTGKGAYLITNDYLAKRDF
 LEMKPLYEWLGLSVSLGFVDIPEYEYAENEKYELYHHDIVYTTNGRLGFDYLDNLADDI
 RAKFLPKLNFAIIDEVDSIILDAAQTPLVISGAPRVQSNLFHKLNSLVFSLSSLTALNWL
 15 STSSDILMPGTSCLKKSALL*

Sequence 513

Contig_0483_pos_4530_3718,
 is similar to (with p-value 3.0e-64)
 20 >sp:sp|P26497|SPOJ_BACSU STAGE 0 SPORULATION PROTEIN J. >pir
 :pir|S18081|A38536_spo0J93 protein - Bacillus subtilis >gp:g
 p|D26185|BAC180K_54 B. subtilis DNA, 180 kilobase region of
 replication origin. NID: g467326. >gp:gp|X62539|BSORIGS_11 B
 .subtilis genes rpmH, rnpA, 50kd, gidA and gidB. NID: g40020
 25 . >gp:gp|Z99124|BSUB0021_201 Bacillus subtilis complete geno
 me (section 21 of 21): from 3999281 to 4214814. NID: g263644
 2.

atgaataatgatgatagtgcaatttattgcactagaattaattagacctaataccttat
 cagccacgtaagacgtttgaagaagaacgactcaatgatttagcttcatcaattcaacaa
 30 catgggtatattacagcctattgtattacgtcaaactgttcaaggttactatattgttg
 ggtgagcgacgatttagagcatctcagttggcgaggattaacagaagtgccagctattatt
 aaagaactatctgatgaagatatgatggaattggcaattattgaaaatttacagagagaa
 gatttaaatgccattgaagaagcagaaagttataaaaaaatgatgacagatttgaatatt
 acacaacaagaggttgcgagacgattaggttaagtcacgtccttatattgccaatatgctt
 35 aggttattacagttacctaataaatgttgctcaaattggttcaacaaggagcgttatcaagt
 gctcatggcgctacgttattaactttgaaagacgccagtaaaaataaaaaagacggcaaaa
 caagccactcaggagtcttgagtgtaaggatttagaggagtagcgtcaatggtttagtc
 agtaaagacatctcaatgaaactggacagagagaccaagggaagtaaacggaaaatgatt
 caacagcaggaaagatttttaaaaaagcaatatggtgcaaaagtagatatttcgacatct
 40 aaaaatgtcgggaaaatcacgtttgaatttaaatctgaagcagaattcaaacgcttgatt
 cgtcaacttaataaagattataaaggaatttaa

Sequence 514

MNNDSDVQFIALELIRPNPYQPRKTFEEERLNDLASSIQQHGIQLPIVLRQTVQGYIIVV
 45 GERRFRASQLAGLTVPAIKELSDDEMMELAIENLQREDLNAIEEAESYKMMTDLNI
 TQQEVARRLGKSRPYIANMLRLQLPKNVAQMVQQGALSSAHGRTLTLTKDASKIKKTAK
 QATQESWSVRYLEEYVNGLVSKDISMKLDRETKGSKPKMIQQQERFLKKQYGAKVDISTS
 KNVGKITFEFKSEAEFKRLIRQLNKDYKEY*

Sequence 515

Contig_0483_pos_3476_2688,
 is similar to (with p-value 4.0e-20)
 >gp:gp|AJ222587|BS16829KB_25 Bacillus subtilis 29kB DNA frag
 ment from ykwC gene to cse15 gene. NID: g2632216. >gp:gp|Z99
 111|BSUB0008_93 Bacillus subtilis complete genome (section 8
 55 of 21): from 1394791 to 1603020. NID: g2633699.
 atgatttttaatttatatcctttagcgattgttggttatagcaattttgaataagattatt
 gagcaagcattttaaattcaaaataaaagcaaaaaaggaataaaaaacgttcaaaaaca
 ctgatttctcttgtacaaaacatagtaaaatatatcgtatggtttgttggttatcacaaca

attttaagtaagtttggtattagcgtcgaaggtatcatcgctagtgtgagttgttaggt
 attgcagttggtttcggtgcgcaacaatagtaaaagatattattacagggtttctttatt
 atctttgaaaatcaatttgatgtgggtgactatgttaaaatcaatagttcaggaactacg
 gtagcagaaggtactgtgaaatctattggtttaagatcaacgcgaattaatacaatttcg
 5 ggagaactgactattttacctaattggttagcatgggggaaattacgaacttttcaattaca
 aatgggactgctattttagaactaccagtatcagttgatgaaaatatagatcaagttgaa
 aagaaactcaatcgtttatttgtttctttacgtagtaaatattacttatttgtcagcgat
 ccagttgttgatggcattgatgcgtagaataataaaggttactatacgaatttcagcg
 gaaacaatttcctggtgaaggattttcaggcgctcgtattattcgtgaaggaagctcaaaaa
 10 atgttttagacaagaaggtattcgcgatgccacaaccagtcatttcaaattataatgaagaa
 aaaagctaa

Sequence 516

MILIYILVAIVVIAILNKIIEQAFKIQNKSKKGNKKRSKTLISLVQNIVKYIWFVVT
 15 ILSKFGISVEGIIASAGVVGIAVGFGAQTIVKDIITGFFIIFENQFDVGDYVKINSSGTT
 VAEGTVKSIGLRSTRINTISGELTILPNGSMGEITNFSITNGTAIVELPVSDENIDQVE
 KKLNLRLFVSLRSKYLLFVSDPVVDGIDAIESNKVTIRISAETIPGEGFSGARIIRKEAQK
 MFRQEGIRMPQPVISNYNEKS*

20 Sequence 517

Contig_0483_pos_2284_1547,
 is similar to (with p-value 2.0e-82)
 >sp:sp|P37518|YYAF_BACSU HYPOTHETICAL 40.1 KD GTP-BINDING PR
 OTEIN IN RPSF-SPO0J INTERGENIC REGION. >gp:gp|D26185|BAC180K
 25 _50 B. subtilis DNA, 180 kilobase region of replication orig
 in. NID: g467326. >gp:gp|Z99124|BSUB0021_197 Bacillus subtil
 is complete genome (section 21 of 21): from 3999281 to 42148
 14. NID: g2636442.
 atggttcaacctaaaaaacaattcctacaacttttgagtttactgatattgcaggtatt
 30 gttaaagggtgcatctaagggcgaaggttttaggaaataaattcctttcacatattcgtgaa
 gtagatgctatatgtcaggtggttcgtgcgtttgacgatgagaatgtaacacatgtatca
 gggcgtgttaattccgcttgatgacatagaagtcattaatatggaacttggttttagcagat
 ttagaatcgttgaaaaacggtttaccgaaaatagagaagatggctcgtcaaaaagataaa
 acagctgagatggaattacgtatatattaacacaaaattaaagaagcgttagaagacggtaaa
 35 ccagtagcgagatttgatttcaatgaggatgatcaaaagtgggttaatacaagctcagtta
 ttaacatctaagaaaatgttatacattgctaattgttggtgaagatgaaattggagataaa
 gataatgataaagtgaagcaattcgtgaatatgcagcaaacgaagattcagaagttatc
 gttattagtgcaaaaatcgaggaagaaatcgctacattagatgatgaagataaagaaatg
 ttcttagaagatttaggcacgaagaaccaggttttagacagactcattagaacaacatat
 40 gatttgatcataaaaaggcgtttgcggtttacaccagaggataggggaaaagttaaggata
 agtttggtgtctatgtga

Sequence 518

MVQPKKTIPTTFEFTDIAGIVKGASKGEGLGKFLSHIREVDAICQVVRAFDDENVTHVS
 45 GRVNPLDDIEVINMELVLADLESVEKRLPKIEKMARQKDKTAEMELRLTQIKEALEDGK
 PVRSIDFNEDDQKVVNQALLTSKKMLYIANVGEDEIGDKDNDKVKAIREYAANEDSEVI
 VISAKIEEEIATLDDDEKEMFLEDLGIEEPGLDRLIRTTYDLIIKRLWRLHQRIGEKLR
 SLVSM*

50 Sequence 519

Contig_0484_pos_4405_3587,
 is similar to (with p-value 3.0e-49)
 >gp:gp|AB001896|AB001896_2 Staphylococcus aureus DNA for sig
 ma70 operon, complete cds. NID: g1943991.
 55 gtgaaggataataatgaagtattaaagttatttatagtttcagattcaattggagaaaca
 gcgcaacggatgattcatgcgacgctgacacagtttccagatttaactcaagtagaaatt
 aagaaatttccatatattaaggacgaacaagaatttttaaatgtcttacaattagctaaa
 gaacagaatgcaattgttgcaacaacattagtgagtgagtcatttaatgcattaggtcat
 cagtttgcaaatgaacatcaaattccctatgtagattacatgtctgagttaattagcata

attaaacaacatacacacgctaaccattaatggaaagtgggtgcgttgcgtaagcttaat
 gatgagtattttaagcgtagaagcaattgagtagtcagtgaatatgatgatggtaag
 cattttacagatattggagaagcggatgctttaatagtaggtgatcacgtacctctaaa
 acgccattaagtatgtacttagctaataaaggatataagattgcaaatattccttttagtc
 5 cctgaagtggctattccagataatgtatttcaacaaaagaatttaaaaggattttggatta
 acagcaagtcaccaattatatcgcaaatatacgacgtaatcgtagcagaaacattagggcta
 tcttcagaatctaattacaatagtttagagcgtatcaaaaaagaattatcttatgctgaa
 gaagtttttagaaaattaaatgcaacggtaattaatacagaatataaatcgatagaggaa
 tcggcattttatattgaaaagtttttagctaaacgttaa

10

Sequence 520

VKDNNEVLKLFIVSDSIGETAQRMHATLTQFPDLTQVEIKKFPYIKDEQEFNLVLQLAK
 EQNAIVATTILVSESNALGHQFANEHQIPYVDYMSELISIIKQHTHAKPLMESGALRKLN
 DEYFKRIEAEISVKYDDGKHFTDIGEADALIVGVSRTSKTPLSMYLANCKYKIANIPLV
 15 PEVAIPDNVFFQKNLKVFLTASPNYIANIRNRRAETLGLSSESNYNLSLERIKKELSYAE
 EVFRKLNATVINTEYKSIEESAFYIEKFLAKR*

Sequence 521

Contig_0484_pos_1484_561,
 putative peptide of unknown function
 20 atgaaaaaattttggggaattttattaattgtgatgtcaattgctcttggggatgttcg
 aatagcaatgattcagatcaatcttctaataagaaagtcacatcaaaaagttcggagaaa
 aaaacggatgtggcgactgaatatacaaaagagaacgaatataaaagaactagaaaaagaa
 gctaaggatcttaacaaaagccagttcttaataagaaatcgatgcacttattacagaaaaa
 25 ggttttacaaacaaaacgggattgcaaggctgggaagactataaaaaattagtgataag
 gtaacacttgagattataaatacacaaaagaatctaagggtcatctatagaagaagtt
 aataagttctttaagataaaaaaggtgtagagattaaacgaatgaaaagtaaggaaaaa
 aatattaagcatatcaattatatgtatgtagatccagatggtaaaaaagcaggtaaaagat
 aagcaacctatgtcctacgctcaaatacttgcaacatttaaaagaaggtaaaattagtagct
 30 acaaatattcaacctggattttttgcttttagacaaaaagaaaatggttaaagctaaagac
 ttagaaaaagtttaagacattggaagatttaacgcgtttgaaagatcctaaagcgacatca
 tatggtattttacagacgaaatataaagggaaccatacactcaagtttcaatattaggc
 agtgattctgatgaagagaatgatatttcctcagccatcttagcttattatctattttca
 ccaacggaattagatagtgacgataatcataaatacgttgaaagttgcatcagcgccattc
 35 ttaagtgtcctaaaacgatttttcatcttatcaactaggcgtatttaaaaaaattatcgaa
 agtagtatgtcggttcgatgaataa

Sequence 522

MKKFWGILLIVMSIALVGCNSNSDSDQSSNEKSSSKSSEKKTVDVATEYTKENEYKELEKE
 40 AKDLKQKPVLEIDALITEKGFTNKTGLQGWEDYKKLVDKVTLADYKYTKESKGSSIEEV
 NKFFKDKKGVIEIKRMKSKEKNIKHINYMYVDPDGKKAGKDKQPMYQAQILATFKEGKLV
 TNIQPGFFALDKKMMVKAKDLEKVKTLLEDLRLKDPKATSYGILQTKYKGPYTQVSILG
 SDSDEENDISSAILAYYLFSPTELDSDDNHXYVEVASAPFLSAQNDFSSYQLGVFKKIIIE
 SSMSFDE*

45

Sequence 523

Contig_0486_pos_255_887,
 is similar to (with p-value 9.0e-20)
 >sp:sp|P33642|YFIT_PSEAE HYPOTHETICAL 39.5 KD OXIDOREDUCTASE
 50 IN FIMT 3'REGION (DADA*) (ORF2). >gp:gp|L48934|PSEPILRV_2 P
 pseudomonas aeruginosa (isolate pRIC351) pilR gene, 3' end of
 cds, dada*, fimT, fimU and pilV genes, complete cds. NID: g
 1161217.
 atgaagtttaagagatattaagcgttatgagtctacagaggtcacttcaatagaacggcat
 55 aatggctattattcagtgaaaaccgatcaatcttcaacaattgaagcgacaaaaattatc
 gttgcaggtggcgcatggtcttcgcaattattaacacaatatcatctacaacgacaagtg
 attggcgttaaagggtgaagttatcttattagaaaaataacgatctttcacttactgagaca
 ttatttatgactaatggtgttacatcggtccaaaacaacccaatcggttttttaattggt
 gcgacgagtgaaatttaataattattctgtcgggtactacagatgaagggtatggattggctt

cttcgccatgcatatcatcgtgtacctaactaaaagacagtcataactgaagaaatgg
 tcaggagtaagaccatacacagaaaaagaaatgccagtcattggatcaaattgatgatggc
 ttatacgtgataagtgggtcattatcgaaacggaatatattgtcacctattatcggtcgt
 gacattgccaatgggtactttctgggtattaaaccatcacgttattcaagttttacagtt
 5 acaaggaggaataatcatgaagtgtatcattaa

Sequence 524

MKLRDIKRYESTEVTSIERHNGYYSVKTDQSSTIEAHKII VAGGAWSSQLLTQYHLQRQV
 10 IGVKGEVILLENNDLSLTETLFMTNGCYIVPKQPNRFLIGATSEFNNSVGGTTDEGMDWL
 LRHAYHRVPQLKDSHILKKWSGVRPYTEKEMPVMDQIDDGLYVISGHYRNGILLSPIIGR
 DIANWLLSGIKPSRYSSFTVTRRNHEVYH*

Sequence 525

Contig_0486_pos_2086_2496,
 15 putative peptide of unknown function
 atggggactacacaagaattaccagtaaaaaacaaaagtttaataagaaaaccattgag
 caaaaagtttttcttatttcgtaaatgataatgggtcaatatttacttgaaaagcgtaaaagaa
 aaacttcttaattgggtatgtggcaatttccaatgagagaacaaacaaatgcaaacgatgtg
 atatctgatgattaggaaaaagtatcgaaacaattaacgaaccagttttaaatataag
 20 catcaatttaccatcttacctgggaaattaaagtatacaatgttacagcacctcttaaat
 ataaaggaataatgatttacctaacaacaaatgacgtgggttaatttagatgataggagcag
 tatatatttcccgtagcaatggataaaatatataagtttattgaaggttaa

Sequence 526

25 MGTQTQELPVKTKSLNKKTIEQKVFLIRNDNGQYLLEKRKEKLLNGMWQFPMREQTNANDV
 ISDDLKXSIETINEPVFKLKHQFTHLTWEIKVYNVTAPLNIKENDLPKQMTWFLNLDREQ
 YIFPVPMDKIYKFIEG*

Sequence 527

30 Contig_0487_pos_6312_5665,
 putative peptide of unknown function
 gtgactaagacagacttatctcatttgcacaacattacaggcattcctctcaatacattg
 tggtagcaaaaaggaacgtggcacatataacgataaattgaagtgcctctttacggacaca
 atgccgagagtgaataaagaacaagagtttaacgaaagagttgtagcaaaagatgaaatt
 35 tggaaagtacagcgagaagtatgacttatacgtaaagcaacttaggcagaatgaaaagacct
 gatggaaaatacaagtttgcgaatgggtgtaaggggtattttcacagttatttataagaat
 aagaagtatcgtgcagcagatattgtgtatgaaacgtttatcggttaacttgaaaaacgga
 ttgcacgcataatccgaaagatagtagatacaacaactttatttcagataaacttattccaa
 40 tctacattacagaaatatagattgtatcgagaaaataaaggtgtatccaaaccagtatatc
 ctagtggatagcgacaacaaaattgtagaagaattcgcaagtacagtagaagctggaaaa
 gtattattcatcgacagacgcaacattgctagaaagtgcaaccgtagatatgtgagtgc
 ggggttgatgtacatgtgggtgatgagtagcagaaggttaaatgcatga

Sequence 528

45 VTKTDLSHLHNITGIPLNTLWYQKERGTYNKLCFFDTMPRVNKKQEFNERVVAKDEI
 WKYSEKYDLYVSNLGRMKRPDGKYKFANGCKGIFTVIYKNKKYRAADIVYETFIGNLKNG
 LHAYPKDSRYNNFISDNLFQSTLQKYRLYRRNKGVSHPVYLVSDNKKIVEEFSTVEAGK
 VLFIDRRNIARKCNRRYVSDGLMYMWADEYEKVN*

Sequence 529

50 Contig_0487_pos_4865_4452,
 putative peptide of unknown function
 atgtggaacgtagaaacgatttatatcgaaagatgaatgggttaaagttaatgacggttcg
 atatacggaaattacaaaggatttagtttagagattatgtattaatgcaatcaacaggctta
 55 aaagataaagaacgggttagagatatagcagggggacatcatcgaaatttgaggatgaatct
 ttttgttatccattcgatgatgaagctatagttgaaacaataaataagagcacaggtaatt
 atagataaaggttaaaggtatttttttggaaaactttatggtaaaaggacagtacgattgct
 aaagaatataaataattattatgatttgccaacatctgaaaaacaatatatttttaagaa
 tgtagtgttgtaggtaattgtgtttgaagatgaaaatttactggaggacagataa

Sequence 530

MWNVETIYIEDEWVKVNDGSIYGITKOLVRDYVLMQSTGLKDKNGVEIYEGDIIEFEDES
 FCYPFDDEAIVETINRAQVIIDKVGIFLENFMVKDSTIAKEYKYYYDLPTSEKTIFFKE
 5 CSVVGNVFEDENLLEDE*

Sequence 531

Contig_0488_pos_7376_8941,

is similar to (with p-value 0.0e+00)

- 10 >sp:sp|P44023|YFCC_HAEIN HYPOTHETICAL PROTEIN HI0594. >pir:p
 ir|E64010|E64010 hypothetical protein HI0594 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32741|U32741_2 Haemophilu
 s influenzae Rd section 56 of 163 of the complete genome. NI
 D: g1573582.
- 15 atgaaaccggtggaacaagcgatcaatgataataaaaaagaaaaaacgttttaactttaga
 atgccaggtgcatttatgattctctttatcctaacagttgtcgcagttatagcaacttgg
 ataatccccgcgggtgcataactcaaaactttcatatgaaccttcacccaagaattaaaa
 attgtcaatcctcatcatcaagtaaaaaagttcctggaacacaaaaggagcttgatcga
 ttaggaggttaaaatcaaaatagaacaatttaaatctggtgcaattaataaaccggttca
 20 attcctaatacttacgaacgtctaaaacaacatccagctggtccttgatcaaattactagt
 agcatggtttaaaggaacctcgaagccgctcgatattatggtctttatacttgttctaggt
 ggactgattggtattgttcaagcgagcgggtccttgaatcaggattgttagcacttact
 caaaaaacgaaaggccacgaatttatgttgattatgttcgtagcaattttaatgattctg
 ggtggaacactatgtggcattgaagaggaagctgtagcgttctatcctgtactcgttcca
 25 atattttattgcgcttgatgatgattctattgtctcagtcggtgcaattttcttagcaagc
 tctgtgggtagtagcattctcaacaatcaaccattctcagtcgctcattgcttctaagca
 gcaggaacaacttttactgatggtctttatttgagaataggcgcttgatcatcgggtgcc
 atatttgttattagttattttattctggtatttgtaaaaaaattaaaaaagatcctaaatcc
 tcttattcttatgaagacaaagcagcatttgaaaaacagtggtctgtgctccatgatgac
 30 ggttcttctgagtttacattacgtaaaaagattattcttacgcttttcgctcctaccattc
 cctattatggtttggggcgctcatgacacaaggatggtggttcccagtcattggcatctgca
 ttcttgatctttaccattgtcatcatgtttattgctggaacaggacaatatggttttaggc
 gaaaaaggcactgtagatgcattcgttaatggcgcttcaagtttagtaggtgatcttta
 atcattggttttagctcgaggaatcaacttagtatgtaataaagggaatgatttctgacaca
 35 atcttgacttttcatcatctatcgtgcaacatatgagtgggcctttatttatcattggt
 ctgctctttatcttttctgtttaggatttatcgtgccgtcctcatcaggatttagcagta
 ctatctatgcctatctttgcgccattagctgatagtaggtataccaagatttggtatt
 gttacaacatatcaattcggtcagtatgcaatgttggttcttagcgccactactggacttgta
 atggcaacacttcaaagttaaacatgcgctactcacactggttacggttctcgatggcct
 40 gttgtcgcgcttggttttaatatttggtggaggcttacttattacacaagttttaataatac
 tcataa

Sequence 532

- 45 MKPLEQAINDNKKKKRNFNRMPGAFFMILFILTVVAVIATWIIIPAGAYSKLSYEPSSQELK
 IVNPHHQVKKVPGTQKELDRLGVKIKIEQFKSGAINKPVSIPTNYERLKHQHPAGLDQITS
 SMVKGTIEAVDIMVFILVLGGLIGIVQASGSFESGLLALTQTKGHEFMLIMFVAILMIL
 GGTLCGIEEEAVAFYPVLVPIFIALGYDSIVSVAIFLASSVGSTFSTINPFSVVIASNA
 AGTTFDGLYWRIGACIIIGAFVISYLFWYCKKIKKDPKSSYSYEDKAAFEKQWSVLHDD
 GSSEFTLRKKIILTLFVLPFIMVWGVMTQGWVFPVMASAFILIFTIVIMFIAGTGQYGLG
 50 EKGTVDAFVNGASSLVGSLIIGLARGINLVNLKGMISDTILHFSSSIVQHMSGPLFIIV
 LLFIFFCLGFIVPSSSGLAVLSMPIFAPLADTVGIPREVIIVTYQFGQYAMLFLAPTGLV
 MATLQMLNMRYSHWLRFVWPVAVFLIFGGGLLITQVLIYS*

Sequence 533

- 55 Contig_0488_pos_9616_9212,

putative peptide of unknown function

atgggaggttatggtgcaatcaaatttgcattaacgcaaagttatcgtttctcaaaagcc
 gctatgctttcagcgccatatgatgtttctatgattggtcaatatcaatggtatgatttt
 actccagaagcgattgtagtaatacgaacatgtcgcgggacatcttttgatccatac

tatttagttgaacaagcaatagacaatggacaaacgttaccacaactatatattacttgt
 ggaactgaagatgaattgtatcaaggtaattgattttgtgaactatttagatgaaaaa
 ggtatttcatatcaatttaaaaaagcgccaggtcatcacgattatgcattttgggataaa
 gcaatagaagatgtcattgaccgttttacatcatcacatatttaa

5

Sequence 534

MGGYGAIKFALTQSYRFSKAAMLSAPYDVSMIGQYQWYDFTPEAIVGNTQHVAGTSFDPY
 YLVEQAIDNGQTLPLQLYITCGTEDELYQGNIDFVNYLDEKGISYQFKKAPGHHDYAFWDK
 AIEDVIDRFTSSHI*

10

Sequence 535

Contig_0488_pos_8631_8224,

is similar to (with p-value 1.0e-35)

>sp:sp|P44023|YFCC_HAEIN HYPOTHETICAL PROTEIN HI0594. >pir:p
 ir|E64010|E64010 hypothetical protein HI0594 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32741|U32741_2 Haemophilu
 s influenzae Rd section 56 of 163 of the complete genome. NI
 D: g1573582.

15

atgataaataaaaggcccactcatatgttgacgatagatgatgaaaagtgcagattgtg
 tcagaaatcattcctttattcaactaagtgtgattcctcgagctaaaccaatgattaaa
 gatacacctactaaactgaagcgccattaacgaatgcatctacagtgcctttttcgctt
 aaaccatattgtcctgttcagcaataaacatgatgacaatggtaaagatcaagaatgca
 gatgccatgactgggaaccaccatccttgtgtcatgacgccccaaaccataatagggaat
 ggtaggacgaaaagcgtaagaataatctttttacgtaatgtaaactcagaagaaccgtca
 tcatggagcacagaccactgtttttcaaagtctgctttgtcttcataa

20

Sequence 536

MINKGPLICTIDDEKCKIVSEIIPLFNTKLI PRAKPMIKDTPTKLEAPLTNASTVPFSP
 KPYCPVPAINMMTMVKIKNADAMTGNHHPCVMTPTIIGNGRKTSVRIIFLRNVNSEEPS
 SWSTDHCFNSNAALSS*

30

Sequence 537

Contig_0488_pos_4443_4015,

is similar to (with p-value 2.0e-50)

>pir:pir|S58181|S58181 fofB protein - Staphylococcus sp. >gp
 :gp|X89875|SSPPDNAFB_1 Staphylococcus sp. plasmidic DNA for
 fosB gene. NID: g927563.

35

atggaaataacaaatgttaatcatatttgtttttcagtgagtgatttaaatacctctata
 caattttataaagatattttacatgggtgacttatttagtatcagatagaacgacagcatat
 ttaactatttggtcactacttggtgactgaatctagaaaaaataaccaagggaatgaa
 ataagtcattcctatacgcacgttgctttctccatagatgaagaagattttcaacagtg
 attcaatggcttaaagagaatcaagtaaatattttaaaagggcgaccaagagacattaaa
 gacaaaaaatcgatatattttacagatctggatgggcataaaattgaattacatactgga
 acattaaaagatagaatggaatattataaatgtgagaagacgcatatgcaattttacgat
 gagttttga

45

Sequence 538

MEITNVNHICFSVSDLNTSIQFYKDILHGDLLVSDRTTAYLTIGHTWIALNLEKNIPRNE
 ISHSYTHVAFSIDEEDFQQWIQWLKENQVNILKGRPRDIKDKKSIYFTDLGCHKIELHTG
 TLKDRMEYYKCEKTHMQFYDEF*

50

Sequence 539

Contig_0488_pos_2775_1777,

is similar to (with p-value 0.0e+00)

>sp:sp|P53557|BIOB_BACSU BIOTIN SYNTHETASE (EC 2.8.1.-). >gp
 :gp|AF008220|AF008220_77 Bacillus subtilis rrnB-dnaB genomic
 region. NID: g2293135. >gp:gp|U51868|BSU51868_5 Bacillus su
 btilis biotin biosynthetic operon genes, complete and partia
 l cds. NID: g1277024. >gp:gp|Z99119|BSUB0016_93 Bacillus sub

55

tilis complete genome (section 16 of 21): from 2997771 to 3213410. NID: g2635411.

atgctaatttttaagaaaaaggagttaaagattatgacattaaacctagctcaacgtgtg
 ttaaatcaagagtcattaacaaaagatgaagcaatatctattttcgaaaatgctgaaatt
 5 gatacatttgatttattaaatgaagcctacacagtgagaaaacattactatggtaaaaaa
 gttaagcttaatatgataattaaatgctaaaagtggtatctgtgagaaagattgtgggtac
 tgtgggcaatctgtaaaaatgaaagaaaagcaacgttatgcacttgttgacaggaccaa
 attaaagaaggcgtcaagtggcaactgaaaatcaaactcggtacatactgtattgttatg
 agtggtagaggtcctagtaacagagaagtcgatcatatttgcgaaacagtagaagatatt
 10 aaaaagatacaccacaactaaagatttgtgctgcttaggattaacgaaagaagaacag
 gctaaaaaattaaaggctgctggtgtcgatcggtataatcataatttaaatacagagtga
 cggtatcacgatgaagtagtaactacacatacatatgaggatagagtgaatacgggtgaa
 atgatgaaagataataatatttctcctgttcagggtgtgatatgtggtatgggagagtcg
 aatgaggacattattgatatggcatttgcctttaagagccatcgatgctgatagcattcct
 15 attaatTTTTTtacatcctattaaaggaactaaatttgggtgattagatttattgtcacca
 atgaaatgtttaagaattatagcgatgttttaggttaataatccaacaaaagaaattcga
 attgcaggtggacgggaggttaaactacggttcattacaaccactcgattgaaagcggct
 aattcaattttttaggagattacttaattacaggcggtaaccgaatgaggaagattat
 cgcatgattgaagatttagggtttgaaatcgacagttaa

Sequence 540
 MLIFKKKELKIMTLNLAQRVLNQESLTKDEAISIFENAEIDTFDLLNEAYTVRKHYGKK
 VKLNMILNAKSGICAEDCGYCGQSVKMKEKQRYALVEQDQIKEGAQVATENQIGTYCIVM
 SGRGPSNREVDHICETVEDIKIHPQLKICACGLTKKEQAKKLKAAGVDRYNHNLNTSE
 25 RYHDEVVTTHTYEDRVNTVEMMKDNNISPCSGVICGMGESNEDIIDMAFALRAIDADSIP
 INFLHPIKGTKEGGLDLLSPMKCLRIIAMFRLINPTKEIRIAGGREVNLRSLQPLALKAA
 NSIFVG DYLI TGGQPNEEDYRMIEDLGFEIDS*

Sequence 541
 30 Contig_0488_pos_1109_369,
 is similar to (with p-value 1.0e-70)
 >sp:sp|P32816|GLDA_BACST GLYCEROL DEHYDROGENASE (EC 1.1.1.6)
 (GLDH). >pir:pir|JQ1474|JQ1474 glycerol dehydrogenase (EC 1
 .1.1.6) - Bacillus stearothermophilus >gp:gp|M65289|BACGLDA_
 35 2 Bacillus stearothermophilus glycerol dehydrogenase (propo
 ed gld) gene, complete cds. NID: gl42976.
 atggatgcaccaacagcagcagtagtctgttatttataacgaagatggatcatttagtggt
 tatgaattctaccctaaaaacctgatacagttatcgtagattctgaaattgttgacaaa
 gcacctgtacgtttatttgcacaggtatgagtgatggtttagcaacattaatcgaagtt
 40 gaattcacattctgtagacaaggcacaacatgttccatggcacaacctacattagcaagt
 ttagcaatcgctcaaaaatgtgaagaggttatttttgaatatggttacagtgttatact
 tctgtagaaaaacatatcgtagacaccacaagtagatgctgtgattgaagccaatacatta
 ctttcagggttttaggatttgaaaacggcggattagcagggtgcacacgcaattcataatgga
 ttcacagctttagaaggggatattccaccacttaactcatggtgaaaaagtggcatacggg
 45 atttttagtacaatttagtacttgaaaatgcgccaaactgaaaaattcatgaaatacaaaa
 ttcttcgataatatcaatatgccaacaacattagaaggtcttcacattgaaaacacaagt
 tatgaagaatttagttcaagtaggtgaacgtgcattaacaccaaatagatacgtttgctaac
 ttaagtataaaaatcactgctgatgaaatcgacagcgaattttaactgttaattgattta
 tctaaaagtcagttcaactaa

Sequence 542
 MDAPTA AVSVIYNEDGSFSGYEFYKPNPDTVI VDSEI VAQAPVRLFASGMSDGLATLIEV
 ESTLRRQGNMFHGKPTLASLAIAQKCEEVIFEYGYSAYTSVEKHIVTPQVDAVIEANTL
 LSGLG FENGGLAGAHAIHNGFTALEDGIHHLTHGEKVAYGILVQLVLENAPTEKFMKYKT
 55 FFDNINMPTTLEGLHIENSTYELVQVGERALTPNDTFANLSDKITADEIADAILTVNDL
 SKSQFN*

Sequence 543
 Contig_0488_pos_0_353,

is similar to (with p-value 5.0e-26)

>pir:pir|S48578|S48578 hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)

5 atgaaaaagttaattcaagataaaaaacacaatttttaaaagatatgcttgatggaattaca
gtttcaacaacgatgttgagttgtatctgacactattgttgtagaaagcataaaaaa
caatcaggtgttgactcgtttctggggcgccagtgacatgaacctgcacacgcagga
ttttagcagaaggcatgctcgatgcagctgtatgtggagaaatcttcacttcacctaca
cctgataaaatattagatgccattaaagctgtggacaatgggtgacggcgttctacttgtt
10 attaaaaactatgcaggagacgttatgaactttgaaatggctcaagaaatggc

Sequence 544

MKKLIQDKNTILKMDLDGITVSNNDVEVVSDTIVVRKHKKQSGVALVSGGGSGHEPAHAG
EVAEGMLDAAVCGEIFTSPDPKILDAIKAVDNGDGVLLVIKNYAGDVMNFEMAQEMA

15 Sequence 545

Contig_0489_pos_142_1740,

is similar to (with p-value 0.0e+00)

>sp:sp|P17894|REC_N_BACSU DNA REPAIR PROTEIN REC_N (RECOMBINAT
ION PROTEIN N). >pir:pir|B35128|B35128 recN homolog - Bacill
20 us subtilis >gp:gp|D84432|BACJH642_227 Bacillus subtilis DNA
, 283 Kb region containing skin element. NID: g2627063. >gp:
gp|M30297|BACREC_N_2 B.subtilis recombination and sporulation
protein (recN, spoIVB) genes, complete cds, arginine hydro
ximate resistance (ahrC) gene, 3' end. NID: g143400. >gp:gp|
25 Z99116|BSUB0013_135 Bacillus subtilis complete genome (secti
on 13 of 21): from 2395261 to 2613730. NID: g2634723.

atgagtgggtgaaactggctcaggaaaatctatcattattgatgccattggacagttaatc
ggtagagagcttcttctgattacgtcagacatgggtgaaaagaaagcaattatcgaaggt
atctttgatatagacgagagtaaaagacgcaattaataactagaatcattagctatagat
30 gttgatgaagattttttattagttaaaagagaaattttcagttctggttaagagtatttgt
cgtattaataaaccaaactgtcactctacaggacttaagaaaagtgatgcaagaactgctt
gatattcatgggtcaacatgaaacgcaatctttacttaagcaaaaatatcatcttcaacta
ttagatgattatgcagacaatcagatttcagatttacttaataatcaactttcttat
aaccaatataaaaaataaacgtaaaagaattagaggaattagaatccgcggaccaggcttta
35 ttacaacgatttagactttaatgaaatttcaatttagaggaactaaccggaagcttcaactgaaa
gaaggcgaagtggaccaacttgaatccgatattaaaagaatttcaaaactccgaaaaatta
aatctagcttttaacaatgcacatcaagtttctaactgatgaaagtgaatacccgatagg
ttgtacgaattaagcaactacttgcacacgattaatgatatcggtccagaaaaattcgta
agattaaaagaggacattgatcaattttactatatactagaagatgcaaaagcatgaaatt
40 tacgacgaaatggctaacactgaattcgatgagcaagttttaaatgagatgaatccaga
atgaatttacttaataatttaaaacgtaaatatggtaaggatattactgaacttattgct
tatcagagtaaaacttgcaaatgaaattgataaaaatagaaaactatgaacaaagtacatca
caattaagggaagaaattaaaacgctttataacgaagtgatagatataggaaaaaaactt
tctcaagaacgtaggcgtgtagcgagagagtttaagggaaccatattgtttctgaaatacaa
45 aatttacaaatgaaagatgctaaccctgaaatttcgtttaaaccattagatgaacctaca
attgaaggtattgaatttgtggaatttttaattagtcctcaaatcggtggaaccacttaaa
agtcttaataaaaatcgcttcaggtgggtgaactttcaagaattatgcttgctctaaaaagt
atatttgttaaatcacgcggccaaaccgcgattctttttgatgaagttgactcgggtgta
tctggtcaagcagcacaaaaaatggctgaaaaaatgcgagatattgctcaatatatacaa
50 gttatttgtattttcacacttacctcaggtagcttcaatgagtgaccatcatcttctaata
agcaaggcatccaatgccgatagaactacaactcaagtcaaagaattgaaagatgaaaac
aaaatagatgaaatagcacgtatgatttcaggagcaagtggtgactgagctcacgagagaa
aatgcaaaagaaatgattaagcaaaatcacaatattta

55 Sequence 546

MSGETGSGKSIIIDAIGQLIGMRASSDYVRHGEKKAIIEGIFDIDESKDAINILES
LAIDVDEDFLLVKREIFSSGKSICRINNQTVTLDLRKVMQELLDIHGQHETQSLLKQYHLQL
LDDYADNQYSDLLNQYQLSYNQYKNRKELEEELESADQALLQRLDLMKFQLEELTEASLK
EGEVDQLESIDIKRIQNSEKLNALNNAHQVLTDESAIPDRLYELSNYLQTINDIVPEKEFV

RLKEDIDQFYMLEDAKHEIYDEMANTEFDEQVLNEYESRMNLLNNLKRKYGKDITELIA
YQSKLANEIDKIENYEQSTSQLREEIKTLYNEVIDIGKKLSQERRRVARELRDHIVSEIQ
NLQMKDANLEISFKPLDEPTIEGIEFVEFLISPNRGEPLKSLNKIASGGELSRIMLALKS
IFVKSRGQTALFDEVD SGVSGQAAQKMAEKMRDIAQYIQVICISHLPQVASMSDHLLI
5 SKASNADRTTQVKELKDENKIDEIARMISGASVTELTRENAKEMIKQNHNI*

Sequence 547

Contig_0489_pos_2000_3421,

is similar to (with p-value 0.0e+00)

10 >sp:sp|P54533|DLD2_BACSU LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4) (DIHYDROLIPOAMIDE DEHYDROGENASE) (LPD-VAL).
atgtcagaaaaaacaatacagatttagtcgtgtaggtggtggtacggcaggatagtagcc
gccatcagagcttctcaattaggaaaaaaagtagcgatagtagaaaaatcactcttaggt
15 ggtacgtgtttacataaaggatgtatacctactaaagcactttttaaatacggctgaagtc
aatcatactattaaaaacgcgcatacatttgggaattgatgtcaatcattttaaataat
ttccctaaaatttttagaacgtaaaagatgctattgttaagcaattgcatgaaggcgtcaat
caactgatgaacatcatcatatagatatttataacggtagtgacgaattatgggaaca
tctatatatttctcctcaaagcggtaacaatttctgtggaatatgaagacggcgaatcagat
20 atactccctaataaaaaatgtgcttataagctactggtcatcaccacagtcctctccggtt
attaaatttgaccataaacaataactatcgagtgatgatcctaaggttaaatacacta
ccacaaagattagcaatcataggtggaggtgttattggtttagaatttgcattctctgatg
aatgatttaggtgctgatgtagtagtaatcgaagcgaatgacagagttcttccctaccgag
agcacacaagttgctgctattgctaaaagaagaattaactaatcgaggcgttacattctac
25 gaaaaattcaattgaccaaagatcattttaaccaaactgataagggtgtaactattaat
atttcagatgagcccggtccaattcgataaagtagtctgttgcaattggtagaaagccta
acaaatgatattgggtttaaataaactcaaaattaagacttctgatgctggtcatattata
acaaatggttatcagcaaactgaagataaacatatatacgcagcaggagattgtataggg
caattacaattggcacacgctcggttcaaaagaagctatagttgcagttgaacatatgttt
30 gattgttctcctatacctatcaattatgacctgataccaaaatgtgtttatacaaaccca
gaaattgcttcaattggtaaaaatttagaacaagcaaaaaagcaggcatcaaagcaaaa
agatcaaaagttccttttaaagctataggaaaggcaataattgaggatgtaacccaatca
aaaggattttgcgagatggtagtttaacaaagatgacgatgaaatcataggtcttaatatg
atagggccacatgttacagaattaataaatgaaatttcattgttacaatttatgaatggc
35 tcatctttagaacttggtttaacaacacatgcacatccttcattatccgaggtagtcag
gaattaggtttaaaagctaattggtcaagcaattcatgtatag

Sequence 548

40 MSEKQYDLVLGGGTAGYVAAIRASQLGKKVAIVEKSLGGLHKGCIPTKALLKSAEV
NHTIKNAHTFGIDVNHFKINFPKILERKDAIVKQLHEGVNQLMKHHHIDIYNGIGRIMGT
SIFSPQSGTISVEYEDGESDILPNKNVLIATGSSPQSLPFKFDHKQILSSDDILRLNTL
PQRLAIIIGGGVIGLEFASLMNDLGADVVEANDRVLPTESTQVASLLKEELTNRGVTFY
ENIQLTKDHFNQTDKGVNTINISDEPVQFDKVLVAIGRKPNNTDIGNNTQIKTSDAGHII
TNGYQQTEDKHIYAAGDCIGQLQLAHVGSKEAIVAVEHMFDCSPIPINYDLIPKCVYTNP
45 ETASIGKNLEQAKKAGIKAKSIKVPFKAIGKAIIEDVTQSKGFCEMVVNKDDDEIIGLNM
IGPHVTELINEISLLQFMNGSSLELGLTTHAHPSLSEVVMELGLKANGQAIHV*

Sequence 549

Contig_0489_pos_3436_4434,

is similar to (with p-value 2.0e-51)

50 >gp:gp|AF012285|AF012285_33 Bacillus subtilis mobA-nprE gene
region. NID: g3282109. >gp:gp|Z99111|BSUB0008_130 Bacillus
subtilis complete genome (section 8 of 21): from 1394791 to
1603020. NID: g2633699.

55 atgatagattacaagtcagcaggccttacagaagaagacctcaaaaaatatataaatgg
atggacttaggaagaaaaacagacgaaaggctatggttactcaatcgtgcaggtaaaatt
ccatttgtgtcagtggtcaaggcaggaagcaactcaaattggtatggcatatgcaatg
caaaaaggtgatatctcatcaccttattatcgtgatttagcatttgcacttatatggga
atttctccattggatactatgttatcagcttttggaaaacgtgatgacattaactcagga

ggtaacaaatgccttctcatttttagtcacaaagaaaaggcattttatctcaaagttct
 ccagtagccactcaaataccacattctgtcggtgctgcattagcacttaaaatggataac
 aagccaaatattgctaccgcaacagttggagaaggcagttcaaatacaaggtagctttcac
 5 gaaggtagtaactttgctgcagttcacaattacctttcgctctgtgtaataattaacaat
 aaatatgcgatatctgtaccagattcactacaatatgctgctgaaaagttatcagatcgt
 gcattagggttacggtagcatggaatacaggttagatggaaatgaccaattgcagtatac
 aaagcgatgaaagaagcaagagaacgagcgctagcaggtgaaggccaacattgatagaa
 gctgtcacttcacgtatgacaccacattcatctgatgatgatgatacatatcgtacaaaa
 gaagaaagagacctattgaaacaagaggattgtaataaaaatttaaaacggccttactc
 10 gatcaaggcatcataaacgaaaattgggttagtcaattggaaaaagagcataaagaactc
 ataatgaagctactaaatctgctgaagcagcaccatattccttcagaagaagaagctttg
 acatatgtttatgaagagggaggtcaacgaaatgactaa

Sequence 550

15 MIDYKSAGLTEEDLKKIYKWMDLGRKTDERLWLLNRAGKIPFVVSQGQEQATQIGMAYAM
 QKGDISSPYRDLAFVTYMGISPLDTMLSAFGKRDDINSGGKQMPSHF\$HKEKGILSQSS
 PVATQIPHVS\$GALALKMDNKPNIATATVGE\$SSNQ\$DFHEGMNFAAVHKL\$PFVCV\$INN
 KYAISVPDSLQYAAEKLSDRALGYGMHGIQVDGNDPIAVYKAMKEARER\$ALAGEGPTLIE
 AVTSRMTPHSSDDDDTYRTKEERDLLQEDCN\$IKFTALLDQGIINENWLSQLEKEH\$KEL
 20 INEATKSAEAPYPSEEEALTYVYEEGGQRND*

Sequence 551

Contig_0489_pos_5558_6742,
 is similar to (with p-value 4.0e-52)
 25 >gp:gp|L25604|BACBMRURBE_4 Bacillus subtilis bmrU, multidrug
 efflux transporter (bmr) and its regulator (bmrR) genes, co
 mplete cds, and branched-chain 2-oxo acid dehydrogenase (bfb
 B) gene, 3' end. NID: g2558636.
 gtgccttcaacaatttctggaacaataacagaattagtggttgaagaaggacaaaactgtc
 30 aatattaacacggtgattttgtaaaatcgattcggaatgggtcaaaatcaaacagaatcg
 gcaaatgagtttaaggaagaacaaaatcagcattctcaatcaaatataaacgtgtcaca
 ttcgaaaataatcctaaaactcatgaaagtgaagtgcatcacgcctctagtcgcgcaaat
 acaatggcagattttcaccagttgtctttaaattagcttctgaacatgatattgattta
 acacaagtcaaaaggaactggttttgaaggtcgtgttactaagaaagatattcaaaatatt
 35 attaacaatccaaacgatcaagaaaaagagaaagaatttaaacaaaacagataaaaaagat
 cattcaacgaaccattgtgactttttacatcaatcctcaactaaaaacgaacactcacca
 ttatcaaatgaacgtgtcgtaccagttaaaggtattagaaaagctatcgacaaaaatag
 gttactagtggtcagcgaaataccacacggttggtgatggttgaagctgatgcaacgaat
 ttggttcagactagaaactatcataaagctcaatttaaacagaatgaggggttacaattta
 40 actttcttgcgttttttgaaaaagctggttcagaggctttaaaagtaaatccattactc
 aatagtagatggcaaggagatgaaattgttatccacaaagatatataatctctattgct
 gttgcagacgatgataagttgtatgtgccagtcattaaaaatgcagatgaaaaatcaatt
 aaaggtatcgcgcgtaaatcaatgatttagctactaaagcaagattagggaaattagca
 caaagtgatatgcaaacggtacattttacggttaataatactggttcttttggttctggt
 45 tcttcaatgggaatcattaatcatccacaagctgccattttacaagtagaatcagtcggt
 aagaaacctgtagttatagatgatattgcaattagaaatatggttaatttgtgtatt
 tcaatcgatcatcgatttctcgatgggtttcaaacgggaaaaatttatgaatcttgttaag
 aaaaaatagaacaatatctattgaaaacacttctatttattaa

Sequence 552

50 VPSTISG\$ITELVVEEGQTVNINTVICKIDSENGQNQTESANEFKEEQNQHSQSNINVSQ
 FENNP\$KTHESEVHTASSRANNGRFS\$PVVFKLASEH\$IDLTQVKGTGFEGRVTKKDIQNI
 INNPNDQ\$EKEFEKQTDK\$KDHSTNHCD\$FLHQ\$STKNEH\$PLSNERV\$PVKGIRKAI\$AQN
 MVT\$VSEIPHGWMMVEADATNLVQTRNYHKAQ\$FKQNEGYNLTFFAFFVKAVAEALKVNPLL
 55 NSTWQ\$GDEIVIH\$KDINISIAVADDDKLYVPV\$IKNADEKSIKGIAREINDLATKARLGKLA
 QSDMQNGTFTVNNTGSF\$GVS\$SMGIINHPQAA\$ILQVESV\$VKKPVVIDDMIAIRNMVNLCI
 SIDHRIL\$GVQ\$TGKFMNLVKKKIEQYSIENTSIY*

Sequence 553

- Contig_0489_pos_6539_6213,
is similar to (with p-value 9.0e-29)
>gp:gp|L25604|BACBMRURBE_4 Bacillus subtilis bmrU, multidrug
efflux transporter (bmr) and its regulator (bmrR) genes, co
mplete cds, and branched-chain 2-oxo acid dehydrogenase (bfm
B) gene, 3' end. NID: g2558636.
atgattccattgaagaaacagaacccaaaagaaccagtattattaaccgtaaagtaccg
ttttgcatatcactttgtgctaattttcctaattctgtcttagtagctaaatcattgatt
tcacgcgcgatacctttaattgatttttcatctgcatttttaatgactggcacatacaac
10 ttatcatcgtctgcaacagcaatagagatattaatatctttgtggataacaatttcattct
ccttgccatgtactattgagtaattggatttacttttaagcctctgcaacagcttttaca
aaaaacgcaaaagaaagttaaattgtaa
- Sequence 554
15 MIPIEETEPKEPVLLTVNVPFCISLCANFPNLALVAKSLISRAIPLIDFSSAFLMTGTYN
LSSSATAIEILISLWITISSPCHVLLSNGFTFKASATAFTKNAKKVKL*
- Sequence 555
Contig_0489_pos_3383_3072,
is similar to (with p-value 6.0e-17)
>sp:sp|P54533|DLD2_BACSU LIPOAMIDE DEHYDROGENASE COMPONENT (E3)
OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
(DIHYDROLIPOAMIDE DEHYDROGENASE) (LPD-VAL).
atgactacctcgataatgaaggatgtgcatgtgtgttaaaccaagttctaaagatgag
25 ccattcataaattgtaacaatgaaatttcatttattaattctgtaacatgtggccctatc
atattaagacctatgatttcacgtcatctttgttaactaccatctcgaaaatcctttt
gattgggttacatcctcaattattgcctttcctatagctttaaaaggaactttgatactt
tttgctttgatgcctgctttttttgcttggttctaaatttttaccattgaagcaatttct
gggtttgtataa
- 30 Sequence 556
MTTSDNEGCACVVKPSSKDEPFINCNNNEISFINSVTCGPILRPMISSSSLLTTISQNP
DWVTSSIIAFPIALKGTLILFALMPAFFACSKFLPIEAISGFV*
- 35 Sequence 557
Contig_0490_pos_4295_4798,
is similar to (with p-value 3.0e-46)
>sp:sp|P42876|UREF_STAXY UREASE ACCESSORY PROTEIN UREF. >gp:
gp|Z35136|SXUREFG_I S.xylosus (C2a) UreF and UreG genes. NID
40 : g511068.
atgagaattgtctaccacgcattaattaacaatgacaaagataaaaatttttagatattaac
caaaaactcttcgtacaaaatctacctaagaaacgcgtattggcgctaagcaaatgggt
acacgcagtgtaaaattagcttttagatctttatgatagtgaatggattcaatggattat
aatcaaatgaaaaacaataaaattaagcttcacctgctgtgtgctttactatgctagga
45 cattttttaggtgtgatgtggaatccatcattgattattatattatcaaaatctct
agccttacccaaaatgcagtaagagcgattcctttaggacaaacagctggacagcaagtc
gtaactgaaatgatagcccatattgagaagacagacatcacataactagaattggacgaa
atcgattttgggtatgactgctcccggttggaacttaataaatggaacatgaaatgtt
catgttcgaatctttatttcataag
- 50 Sequence 558
MRIVYHALINNDKDKILDINQKLFVQNLPKETRIGAKQMGRMVKLALDLYDSEWIQWYY
NQMKNNKIKLHPAVCFMTLGHFLGVDVESIIDYYLYQNISLTONAVRAIPLGQTAGQQV
VTEMLAHIEKTRHHILELDEIDFGMTAPGLELNQMEHENVHVRIFIS*
- 55 Sequence 559
Contig_0490_pos_4880_5425,
is similar to (with p-value 5.0e-91)
>sp:sp|P42877|UREG_STAXY UREASE ACCESSORY PROTEIN UREG. >gp:

- gp|235136|SXUREFG_2 *S.xylosus* (C2a) UreF and UreG genes. NID : g511068.
- gtgggttaaacgccttgcgaaaaaaatgagtattggcggttattactaatgatattctataact
 aaagaagatgaaaaaatactagtttaatacaggtgttttaccagaagatagaattatcggg
 5 gtggaactggaggtgtcctcatcacagctattcgtgaagacgcctcaatgaacttcgca
 gccatagatgaattattagaacgtaatgatgatattgaacttatttttattgaatcaggt
 ggcgataacttagcggctacttttagtccagaactcgttgacttttcaatttatatcatt
 gatgttgctcaggcgcaaaagattccacgtaaaaggtggacaaggtatgattaaatctgat
 ttcttcattattaataaaactgaccttgccacatattgtgggtgcttcattagatcaaagt
 10 gctaaagatactgaagtatttcgtggaaatcatccattcgtttttacaaatttaaaaact
 gatgaaggtttagaaaaagttattgagtggttgagcagcagcgtcttactgaaagggtta
 acttaa
- Sequence 560
- 15 VVKRLAKKMSIGVITNDIYTKEDKILVNTGVLPELDRIIGVETGGCPHTAIREDA SMNFA
 AIDELLERNDDIELIFIESGGONLAATFSPVLVDFSIYIIDVAQGEKIPRKGQGMKSD
 FFIINKTDLAPYVGASLDQMAKDTEVFRGNHPFAFTNLKTDEGLEKVIEWIEHDVLLKGL
 T*
- 20 Sequence 561
 Contig_0490_pos_5557_0,
 is similar to (with p-value 7.0e-46)
 >sp:sp|Q07400|URED_BACSB UREASE ACCESSORY PROTEIN URED. >pir
 :pir|G36950|G36950 ureD protein - *Bacillus* sp. (strain TB-90
 25) >gp:gp|D14439|BACUREA_7 Thermophilic *Bacillus* genes for ur
 ease subunits and urease accessory proteins, complete cds. N
 ID: g393296.
- gtgccaactttctatattgtcaatgtgggtggaggttatctagatggagatagataaccgt
 gtcaatgtcaacttagaagataatgcacaagtgcgcttacttctcaagggtgcaactaaa
 30 atatatataaacgcctaataatgaccatgtagaacagtatcaaactgttaatttatcaaataca
 tcgtatatggaattttagtagcagatcctattattgcctatgaaaacgctaaatttttccaa
 cataatacgtttaatcttaagaagatagtgctatattttacacagatatattgactccc
 ggctattcatctaattggccaagatttcacgtataattatatgcatcttactaatgaaatt
 tacattgacaatacaattagttgttttcgataaacatgatgttaagtcctgataaaagccga
 35 cttgacggtattgggtatatggaaaattatacacacttaggatcagcttattttatcat
 ccagatgtaaaaccaaagtttcatagacgatatttacgtggcggttgctgattttcaaaaa
 caatacagctgtagaataaggtatctcacaattacactcatggattggccggttcgtatt
 ttgactaaaagaactcaaataatagaagaattttgactcgtgttcaatcat
- 40 Sequence 562
 VPTFYIVNVGGGYLDGDRYRVNVNLEDNAQVTLTSQGATKIYKTPNDHVEQYQTFNLSNQ
 SYMEFVADPIIAYENAKFFQHNTFNLKEDSAIFYTDILTGPYSSNGQDFTYNYMHLTNEI
 YIDNQLVVDNMMLSPDKSRLDGIGYMENYTHLGSAYFIHPDVNQSFIDDIYVAVADFQK
 QYDCRIGISQLPTHGLAVRILTKRTQIIIEILTRVQ SX
- 45 Sequence 563
 Contig_0490_pos_3469_3101,
 is similar to (with p-value 3.0e-34)
 >sp:sp|P02395|RL7 MICLU 50S RIBOSOMAL PROTEIN L7/L12 (MA1/MA
 50 2). >pir:pir|A02771|R7MCML ribosomal protein L7/L12 - *Microc
 occus luteus*
- atggctaatacaagaacaaatcattgaagcaattaaagaaatgtcagttattagaattaac
 gatttagtaaaagcaattgaagaagaatttggtgtaactgcagcagctccagtagcagca
 gcaggtgcagctggtggcgagatgcagcagctgaaaaaactgaatttgatgttgaaata
 55 acttcagctggatcttcaaaaattaaagttgttaaaagcagttaaagaagcaactggctta
 ggattaaaagatgctaaagaattagtagatggagctcctaaagtaattaaagaagctatg
 cctaaagaagatgctgaaaaacttaagaacaattagaagaagttggagctagcgtagaa
 ttaaaatag

Sequence 564

MANQEQIIIEAIKEMSVLELNDLVKAIIEEFVGTAAAPVAAAGAAGGGDAAAEKTEFDVEL
TSAGSSKIKVVKAVKEATGLGLKDAKELVDGAPKVIKEAMPKEDAEKLKEQLEEVGASVE
LK*

5

Sequence 565

Contig_0490_pos_3035_2301,

is similar to (with p-value 2.0e-81)

>pir:pir|S59955|S59955 hypothetical protein 202 - Staphyloco-
ccus aureus >gp:gp|X64172|SARPLRPO_2 S.aureus rplL, orf202,
10 rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypot-
hetical protein ORF202, DNA-directed RNA polymerase beta & b-
eta' chains. NID: g677848.

atgcaaatacaacctgaattctttgagaagttaaaaccccggtattccaataacgggggt
15 tttcaattaattaatacaagcactactaaatgcttttgattgaatgataaaaatagaggt
gaagaaatgagtcattattatgatgaacaacctgatgttaaaagtaacccaaaaagaatt
agttatcaaattaaaaatgcgcaactagagcttactactgatgctggagttttttcaaaa
gataatgtagattttggatctgacttactaattaaaacttttttaaaagaacatcctcca
ggcccaagtaaaaccatcgcgatgtaggatgtggatattggtcctatcggttttagcaata
20 ggaaaagtatctccacaccatcaaatcaccaatgttgatattaacaatagagccttgccg
ttggcagaaatgaataagacgaaaaatcaagtggataatgtaacgattatagaaagcgat
tggttatctgctgcaatcatcagtgctttgattacattttaactaatccccctattaga
gctggtaaggacattgttcacgaatctttgaacaagcggttgacagactcaagactacg
ggtgaactttatgtcgtcattcaaaaaagcaaggtatgccttcagctaaaaagaaaata
25 gaagaactatttgcaatgtagaaattatagctaagagtaaaaggatattatattttgaaa
agtataaaaggttga

Sequence 566

MQIQPEFFEKLPFRYSNNGVFQILINTSILNAFGLNDKNRGEEMSHYYDEQPDVKSNNPKRI
30 SYQIKNAQLELTDDAGVFSKDNVDFGSDLLIKTFLKEHPPGPSKTIADVCGYGPGLAI
GKVSPHHQITMLDINNRLALAEMNKTKNQVDNVTIIESDCLSAVNHCQFDYILTNPPIR
AGKDIVHRIFEQAFDRLKTTGELYVVIQKKQGMPSAKKKIEELFGNVEIIAKSKGYIILK
SIKG*

35 Sequence 567

Contig_0490_pos_0_1944,

is similar to (with p-value 0.0e+00)

>sp:sp|P47768|RPOB_STAAU DNA-DIRECTED RNA POLYMERASE BETA CH-
AIN (EC 2.7.7.6) (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE
40 BETA SUBUNIT). >pir:pir|S59951|S59951 DNA-directed RNA polym-
erases (EC 2.7.7.6) beta chain - Staphylococcus aureus >gp:gp-
|X64172|SARPLRPO_3 S.aureus rplL, orf202, rpoB(rif) and rpoC
genes for ribosomal protein L7/L12, hypothetical protein OR-
F202, DNA-directed RNA polymerase beta & beta' chains. NID:
45 g677848.

atgttcagagatatttctccaattgaagatttctactggcaacctatctttagaatttgta
gattacagattaggtgaacctaaatgatgatttagaagaatcaaaaaaccgtgacgctact
tatgctgcacctcttcgtgtgaaagtgctcttattattaaagaaactgggtgaagttaaa
gaacaagaagtcttcattgggtgattttccactgatgacagatacaggtacgttcgtaatt
50 aatgggtgctgaacgtgttatcgtatctcaattagttcgttcaccatccgtgtattttaac
gagaaaatcgataaaaatggacgtgaaaactatgatgcaacaatcattcctaaccgtggt
gcttggttagagtatgagacagatgctaaagatgttgatatgttcgtatcgatagaaca
cgtaaattaccattgactgtattactacgtgcgctagggtttctcaactgatcaagaaatt
gttgacttattaggagacagcgaatatttacgtaatacattagaaaaagatgggacagaa
aatacagaacaggctttatttagagatttatgaacgtttacgtcctggcggaaccaccaaca
55 gtagaaaatgctaaaagtttatattattctcgtttcttcgaccctaaccgctatgattta
gccagtgtaggctggttataaaagcgaacaaaaaattacacctaaccatcggttggtcaat
caaaaattagcagaaccaattgttaacagtgaaactgggtgaaattgttggtgacgaagga
acagtggttagatcgctcgtgaaacttgacgaaatcatggacgtattagaaacaaacgctaatt

agcgaagtatttgaacttgaaggtagcgtaattgacgaacctgtagaaatccaatctatt
 aaagtgtatgtgcctaacgatgaagaaggtagtgcactactgtcattggtaatgcatta
 cctgattctgaagttaaagtattactccagcagatattgttgcctcaatgagttatttc
 5 tccaacttattgaatggcatttggttatcacagatgatattgatcatctaggtaatcgctcg
 ttacgttctgtcggtgagctattacaaaatcaattccgtatcggtttatccagaatggaa
 cgtgttgcgtgaaagaatgtcaatacagatacagattctattacgccacaacaactc
 attaatatcagaccagttattgcatcaatcaaagaattcttggtagttcacaattatct
 caattcatggaccaagctaaccggttagcagagttaacgcacaaacgtcggttatctgct
 ctagggcctgggtgattaacacgtgaacgtgctcaaattggaagtgcgtgacgttactac
 10 tctcactatgggctgatgtgtccaattgaaacacctgagggctcctaatttggtttaata
 aactcattgtcaagttatgctagagtgaattggttttattgaaacccatctcgt
 aaagtggatttagatacaaaactcaatcactgatcaaatagattatttgacagctgatgaa
 gaggatagttacgttgttgcacaggctaattctagacttgatgaaaatggctgtttctta
 gatgatgaagttgttgtgtcttccgttgtaataacactgttatggctaaagaaaaaatg
 15 gattacatggacgtatcaccaaaacaagttgtttcagcagcaacagcatgtattccattc
 ttagaaaatgacgactctaaccgtgcttaattgggagcaaacatgcaacgtcaagcgggtg
 cctttaatgaatccggaagctccatttgtgggtacaggtatggaacacgtatccgcaaga
 gactctgggtgctgcaattactgctaagcatagaggacgcgttgagcatgttgaatcta
 gaaatttttagttcgtcggttagtc

20 Sequence 568
 MFRDISPIEDFTGNLSLEFVDYRLGEPKYDLEESKNRDATYAAPLRVKVRLIIKETGEVK
 EQEVFMGDFPLMTDTGTFVINGAERVIVSQLVRSPSVYFNEKIDKNGRENYDATIIPNRG
 AWLEYETDAKDVVYRIDRTRKLPITVLLRALGFSTDQEIVDLLGDSEYLRNTLEKDGTE
 25 NTEQALLEIYERLRPGEPPTVENAKSLLYSRFFDPKRYDLASVGRYKANKKLHLKHLRFN
 QKLAEPVIVNSETGEIVVDEGTVLDRRLDEIMDVLETNANSEVFEELEGSVIDEPVEIQSI
 KVVYPNDEEGRTTVIGNALPDSEVKCITPADIVASMSYFFNLLNGIGYTDIDHGLNRR
 LRSVGELLQONQFRIGLSRMERVVRERMSIQDTSITPQQLINIRPVIAKIEFFGSSQLS
 QFMDQANPLAELTHKRRLSALGPGGLTRERAQMEVRDVHYSHYGRMCPIETPEGPNIGLI
 30 NSLSSYARVNEFGFIETPYRKVDLDTNSITDQIDYLTADEEDSYVVAQANSRLDENGRL
 DDEVVCRFRGNNTYMAKEKMDYMDVSPKQVVSAAATACIPFLENDSDNRALMGANMQRQAV
 PLMNPEAPFVGTGMEHVSARDSGAAITAKHRGRVEHVESNEILVRRLV

Sequence 569
 35 Contig_0491_pos_1640_0,
 putative peptide of unknown function
 gtggatgatgtgacaaaatattggtccagttgatggagatccgatcacgtcaacggaagaa
 attccattcgacaagaaacgtgaattcaatcctgattttaaaccaggtgaagagcgtgtt
 aaacaaaaagggtgaaccaggaacaaaaacaattacaacaccaacaactaagaaccatta
 40 acaggggaaaaagtggcgaagggtgaaccaacagaaaaataacaaaacaaccagtagat
 gaaatcacagaattatgggtggcgaagaaatcaagccaggccataaggatgaatttgatcca
 aatgcaccgaaaggtagccaagaggacgttccaggtaaaccaggagttaaaaaccctgat
 acaggcgaagtagtcacaccaccagtggtgatgtgacaaaatattggtccagttgatgga
 gatccgatcacgtcaacggaagaaattccattcgacaagaaacgtgaattcaatcctgat
 45 ttaaaccaggtaaagagcgctttaaacaagaaagggtgaaccaggaacaaaaacaattaca
 acaccaacaactaagaaccattaacaggggaaaaagtggcgaagggtgaaccaacagaa
 aaagtaacaaaaacaaccagtagatgaaatcacagaatatgggtggcgaagaaatcaagcca
 ggccataaggatgaatttgatccaaatgcaccgaaaggtagccaagaggacgttccaggt
 aaaccaggagttaaaaatcctgatacaggcgaagtagttactccaccagtggtgatgtg
 50 acaaaatattggtccagttgatggagatccgattacgtcaacggaagaaattccgtttgat
 aaaaaacgcgaatttgatccaaacttagcgccagggtacagagaaagtcgttcaaaaagggt
 gaaccaggaacaaaaacaattacaacaccaacaactaagaaccattaacagggggaaaaa
 gttggcgaagggtgaaccaacagaaaaagtaacaaaaacaaccagtggtgaaatcgttcat
 tatgggtggcgaagaaatcaagccaggccataaggatgaatttgatccaaatgcaccgaaa
 55 ggtgacaaagagagcgttccaggtaaaccaggagttaaaaaccctgatcacaggcgaagta
 gttactccaccagtggtgatgtgacaaaatattggtccagttgatggagatccgattacg
 tcaacggaagaaattccgtttgataaaaaacgcgaatttgatccaaacttagcgccagggt
 acagagaaagtcgttcaaaaagggtgaaccaggaacaaaaacaattacaacaccaacaact
 aagaaccatttaacagggggaaaaagtggcgaagggtgaaccaacagaaaaagtaacaaaa

caaccagtggatgaaatcggtcattatgggtggcgaagaaatcaagccaggccataaggat
 gaatttgatccaaatgcaccgaaaggtagtcaaacaacgcaaccaggtgaagccgggggtt
 aaaaatcctgatacaggcgaaagtagttactccacctgtggatgatgtgacaaaatatggt
 ccagttgatggagatccgatcacgtcaacggaagaaattccattcgacaagaaacgtgaa
 5 ttcaatcctgatttaaaaccaggtgaagagcgtgttaacaaaaaggtgaaccaggaaca
 aaaacaattacaacaccaacaactaagaaccattaacaggggaaaaagttggcgaaggt
 gaaccaacagaaaaataacaaaaacaaccagtagatgaaatcacagaatatggtggcgaa
 gaaatcaagccaggccataaggatgaatttgatccaaatgcaccgaaaggtagccaagag
 gacgttccaggtaaacaggagttaaaaaccctgatacaggcgaaagtagtcacaccacca
 10 gtggatgatgtgacaaaatatggtccagttgatggagatccgatcacgtcaacggaagaa
 attccattcgacaagaaacgtgaattcaatcctgatttaaaaccaggtaaagagcgcgtt
 aaacagaaaggtgaaccaggaacaaaaacaattacaacaccaacaactaagaaccatta
 acaggggaaaaagttggcgaaggtgaaccaacagaaaaagtaaca

15 Sequence 570

VDDVTKYGPVDGDPITSTEEIPFDKKREFNPDLKPGEERVKQKGEPGTKTITPTTKNPL
 TGEKVGEGETEKITKQPVDEITEYGGEEIKPGHKDEFDPNAPKGSQEDVPGKPGVKNPD
 TGEVVTTPVDDVTKYGPVDGDPITSTEEIPFDKKREFNPDLKPGKERVKQKGEPGTKTIT
 TPTTKNPLTGEKVGEGETEKITKQPVDEITEYGGEEIKPGHKDEFDPNAPKGSQEDVPG
 20 KPGVKNPDTEGVVTPPVDDVTKYGPVDGDPITSTEEIPFDKKREFDNLAPGTEKVVQKG
 EPGTKTITPTTKNPLTGEKVGEGETEKITKQPVDEIVHYGGEEIKPGHKDEFDPNAPK
 GSQEDVPGKPGVKNPDTEGVVTPPVDDVTKYGPVDGDPITSTEEIPFDKKREFDNLAPG
 TEKVVQKGEPGTKTITPTTKNPLTGEKVGEGETEKITKQPVDEIVHYGGEEIKPGHKD
 EFDPNAPKGSQTTQPGKPGVKNPDTEGVVTPPVDDVTKYGPVDGDPITSTEEIPFDKKRE
 25 FNPDLKPGEERVKQKGEPGTKTITPTTKNPLTGEKVGEGETEKITKQPVDEITEYGGEE
 EIKPGHKDEFDPNAPKGSQEDVPGKPGVKNPDTEGVVTPPVDDVTKYGPVDGDPITSTEE
 IPFDKKREFNPDLKPGKERVKQKGEPGTKTITPTTKNPLTGEKVGEGETEKITKQPVDEITEYGGEE

Sequence 571

30 Contig_0491_pos_3423_3109,
 putative peptide of unknown function
 gtgatttcatctactggttgttttgttattttttctgttggttcaccttcgccaactttt
 tccccgttaaatgggttcttagttgttggtgttgtaattgttttgttcctgggtcacct
 ttttgtttaacacgctcttccactgggttttaaatcaggattgaattcacgtttcttgtcg
 35 aatggaatttcttcggttgacgtgatcggtatccatcaactggaccatattttgtcaca
 tcatccacaggtggagtaactacttcgcctgtatcaggatttttaacccccggccttacct
 ggttgctgtgtttga

Sequence 572

40 VISSTGCFVIFSVGSPSPTFSPVNGFLVVGVIIVFVPGSPFCLTRSSPGFKSGLNSRFLS
 NGISSVDVIGSPSTGPYFVTSSTGGVTTSPVSGFLTPGLPGCVV*

Sequence 573

45 Contig_0491_pos_2271_1351,
 putative peptide of unknown function
 gtgatttcatctactggttgttttgttactttttctgttggttcaccttcgccaactttt
 tccccgttaaatgggttcttagttgttggtgttgtaattgttttgttcctgggtcacct
 ttctgtttaacgcgctcttccactgggttttaaatcaggattgaattcacgtttcttgtcg
 aatggaatttcttcggttgacgtgatcggtatccatcaactggaccatattttgtcaca
 50 tcatccactggtggtgtgactacttcgcctgtatcagggttttaactcctgggttacct
 ggaacgtcctcttggttacctttcggtgcatttgatcaaatcatccttatggcctggc
 ttgatttcttcgccaccatattctgtgatttcatctactggttgttttgttatttttct
 gttggttcaccttcgccaacttttccccgttaaatgggttcttagttgttggtgttgta
 attgttttgttcctgggttcaccttttgtttaacacgctcttccactgggttttaaatca
 55 ggattgaattcacgtttcttgtcgaatggaatttcttcggttgacgtgatcggtatcca
 tcaactggaccatattttgtcacatcatccacaggtggagtaactacttcgcctgtatca
 ggatttttaacccccggccttacctgggttgctgtttgactaccttcggtgcatttgga
 atcaaatcatccttatggcctggccttgatttcttcgccaccataatgaacgatttcac
 cactggttggtttgttatttttctgttggttcaccttcgccaacttttctcctgtatt

aggattgacataagttgggtgttgttgttgtttcaattcctgggtcacctttttggactac
tttttctgtacctggggctaa

Sequence 574

5 VISSTGCFVTFVSVGSPSPTFSPVNGFLVVGVVIVFVPGSPFCLTRSLPGFKSGLNSRFLS
NGISSVDVIGSPSTGPYFVTSSSTGGVTTSPVSGFLTPGLPGTSSWLPGAFGSNSSLWPG
LISSPPYVISSTGCFVIFVSVGSPSPTFSPVNGFLVVGVVIVFVPGSPFCLTRSSPGFKS
GLNSRFLSNGISSVDVIGSPSTGPYFVTSSSTGGVTTSPVSGFLTPRLTWLRCLTTFRCIW
10 IKFILMAWLDEFATIMNDFIHWLFCYFFCWFTFANFFSCIRIDISWCCCCFNSWFTFLDY
FFCTWG*

Sequence 575

Contig_0491_pos_518_78,
putative peptide of unknown function

15 atgacgtacttgccgtattggagtcacttacgctcgttgatgttgacgcacttttctgaag
tcgatgtactcgttgatcctgaaagactcgtacttgctcgagccacttaatgacgtacttg
tactgtttgattcactcgttgatgcagatgcgctatcagacatcgacgtactcgtgatt
ctgataacttcttacttgactcgtgattcactctcactcgttgatgtggatgcacttt
ctgatgtcgacgtgcttgttgaaactgaaacgcttgctgcttgctgactcacttaagatg
20 tgcttgcaactgtttgagtcgctcacactgttgacgttgacgcactgtctgatgtcgatg
tactcgttgaaatccgaaatgcttgacttgctcgagtcacttaaggacgtacttgcaactgt
ttgagtagcttacactcatag

Sequence 576

25 MTYLRYSHLRSLMLTHFLKSMYSLILKDSYLSHLMTYLYCLIHSLMQMRYQTSTYSLI
LITSYLYSLIHSHSLMWMHFLMSTCLLNLKRLCLSTHLKMLHCLSRSHLLTLTHCLMSM
YSLNPKCLYLSSHRLTYLHCLSSLHS*

Sequence 577

30 Contig_0493_pos_2737_3663,
is similar to (with p-value 0.0e+00)
>gp:gp|Z99108|BSUB0005_72 Bacillus subtilis complete genome
(section 5 of 21): from 802821 to 1011250. NID: g2633055. >g
p:gp|D78509|D78509_8 Bacillus subtilis YfjG-YfjR genes, comp
lete cds. NID: g2780390.

35 atggaagacgtgacagatattgtctttcggcatgttgatcagtgaaagctgagaccagat
gtattttttactgaattttaccaatactgagagttactgtcaccctgaaggtattcatagt
gtgcgcggacgcttaacttttagtgacgacgaacaaccaatggtagcgacacatctggggc
gataaaccagaacaattccgagaaatgagatcggcttagcggatattgggttttaaaggat
40 atagatttaaatatgggttgccctgtcgcaaacggttgcaaaaaaggtaaggatccggc
ttaattctacgacctgaaacggcagccgaaatcattcaagcttctaaagcaggtggtcta
ccggtcagtgtaaaaaacacgcttaggttattacgatatcgatgaatggcgagactggtta
aaacacgtcttcgaacaagatcgcaaatattatccattcatctacgtacccgtaagag
atgagtaaaagtagatgcacactgggaattaatcgaagcaatcaagacattacgtgatgaa
45 attgcgccaatacactattaactatcaatgggtgatatccccgatagacaaactgggtcta
gaactcgcaataaataatgggtattgatggcattatgattggtagaggatcttcataaac
ccattcgcatattgaaaaggaaacgcgcaaacattcaagcaagaattatttaggtttatta
cgcttacatctctcttttattgaaaaatatgataaagatgaagcccgcacactcaaaagt
ttacgcagattcttcaaaatctacgtacggcattagaggcgctagcgaactccgcat
50 caattaatgaacaccaatccattgccgaagcaagagaactactcgatacttttgaagca
cgtatggatgcacgttcagaagtataa

Sequence 578

55 MEDVTDIVFRHVSEAAARPDVFFTEFTNTESYCHPEGIHSVRGRLTFSDDDEQPMVAHIWG
DKPEQFREMSIGLADMGFKGIDLNMGCPVANVAKKGKSGSLILRPETA AEI IQASKAGGL
PVSVKTRLGYYDIDEDWDLKHVFEQDIANLSIHLRTRKEMSKVDAHWELEIAIKTLRDE
IAPNTLTITINGDIPDRQTGLELANKYGIDGIMIGRGIFHNPF AFEKEPREHSSKELLGLL
RLHLSLF EKYDKDEARHFKSLRRFFKIYVRGIRGASELRHQLMNTQSI AEARELLDTFEA
RMDARSEV*

Sequence 579

Contig_0493_pos_5647_6057,

putative peptide of unknown function

5 atgtcaaaaaagaaaatcttgatctttatttagtgatattaatcatttttgggggcttt
 tatctcaaaatgaaatataaacgaaaaagaaaaacagaaagaaatctactacaaagagcaa
 caagaacgtatcacgctttatcttaaatacaacactaaagaacctaatatcatcaaatct
 gtccatttcacaagtttaaaacagggaaccaatgggtgacgctgttattgaaggctatatc
 aataacaataaaaaagatgattttgttgcatctgacacctgaaaacaattatcaattt
 10 ggaggcagacttatagcagacgttaaaatatttaattacttaaacgggctaataatgaatct
 aaatcacccgatgaaatcaaaaaagatttagacaaaaagaaagaacactaa

Sequence 580

MSKKKILIFISVILIIFFGGFYLMKYNEKEKQKEIYYKEQQERITLYLKYNTPKEPNIKS
 15 VHFTSLKQGPMDAVIEGYINNNKKDDEFAFASPENNYQFGRLLIADVKIFKLLKPANES
 KSPDEIKKDLDDKKKEH*

Sequence 581

Contig_0493_pos_8145_8723,

putative peptide of unknown function

20 atgctaggatttgcagggggattgggatacagtcattataaagattcaaaatcgaacact
 gatgtagcttcaaaagagactcagacttccaataaaaaacactcatgaagatacaacttca
 caaggtaaaatgcaaaatcaagttaatagccaaacaaacgaagtatcaaatgggacatca
 actaaaacacttagtgaaaaagcaaaagcagttaagagaagcttttaacgtcaatgatgag
 25 gaagctcaaatttttagcagatgaaatcgatagagcagatgtaataaagatggcacgatt
 acaacgggatgaaatgacgcctacttttagatcgctttacaaaagaagggaattccaacca
 tctgctgggtggtacaaactagcgaaacacctcaccctaataatatacagcagaagatgctaga
 catatgtctgatgatgaattttctagacgcgtatacagaaggcatgtcagatgatgaagct
 gctactattcacgaaagtgtcaagaatctaacgagtatatgaaatttttaagaggacaa
 30 gttgaagcacgtgcaaaaggacagggcggaattatttaa

Sequence 582

MLGFAGGLGYSHYKDSKSNSTDVASKETQTSNKNTHEDTTSQGKMNQVNSQTNEVSNGTS
 TKTLSEKARQLREAFNVNDEEAQILADEIDRADVNKDGTTTDEMTPTLDRFTKEGKFQP
 35 SAGGTTSETPHPKYTAEDARHMSDDEFLDAYTEGMSDDEAATIHESAQESNEYMKFLRGQ
 VEARKGQGGNY*

Sequence 583

Contig_0493_pos_12191_11406,

putative peptide of unknown function

40 atgcaacattcaagcaaaataatagatatttgaagtttcttaattttaacgatttttatt
 ggaggatgtggttttataaaataaagaagatagcaaaagaaacggaaatcaaacaaaacttt
 aataaaatgttagacgtgtatccaactaaaaatctagaagacttttatgataaagagggc
 tatcgtgatgaagagtttgataaagatgacaaaggaacatggattattaggtctgaaatg
 45 acaaaacagccaaaaggtaaaattatgacctcaagagggtatggttctctatatcaatcgc
 aacactagaacagccaaaagggtattttttaatagataagataaaagatgatagtaatggt
 agaccgatagagaatgaaaagaaataccctgtaaaaatgaaccataataagatctttcca
 acaaagccaatatctgatgataagttaaaaaaaagaaattgaaaacttcaaattttttgtg
 caatatggaaatttttaaaaacttaaggattataaaaaacggggatattttatacaatcct
 50 aatgttcctagttatttctgcgaaatatcaattgagtaataatgaatataacgtacaacaa
 ttaagaaaaagatatgacatcccaactaaaaaagcacctaaactattgttaaaaggggat
 ggcgacttaaaaggatcatccgtaggtcatagagacctagaatttacctttgtagagaat
 aagaaagaaaacatcttttttacggatagatttaatttttaaacgactgagcgtgatgaa
 tcatga
 55

Sequence 584

MQHSSKIIVFVSFLILTIFFGGCGFINKEDSKETEIKQNFNKMLDVYPTKNLEDFYDKEG
 YRDEEFDKDDKGTWIIIRSEMTKQPKGKIMTSRGMVLYINRNRTAKGYFLIDKIKDSSNG
 RPIENEEKYPVKMHNHNIFFTPKIPISDDKLKKEIENFKFFVQYGNFKNLKDYNKDILYNP

NVPSYSAKYQLSNNEYNVQQLRKRYDIPTKKAPKLLKGDGLKGSSVGHROLEFTFVEN
KKENIFFTDSINFKPTERDES*

Sequence 585

5 Contig_0493_pos_11134_10166,
is similar to (with p-value 2.0e-20)
>gp:gp|AJ222587|BS16829KB_25 Bacillus subtilis 29kB DNA frag
ment from ykWc gene to cse15 gene. NID: g2632216. >gp:gp|Z99
111|BSUB0008_93 Bacillus subtilis complete genome (section 8
10 of 21): from 1394791 to 1603020. NID: g2633699.
atgacttactgctcattatctataaaagatttatacattgaaaatattaattgttttaaca
ataggaggtcatgttgttattatgagtcagtttaaggacacattatataaaactatttgag
ccaatgatgaaaatagagtttctatcaaaatcttttggttaattcttttaattatacttgct
tatactcttgatgggtatgattgtaattgcgatatcaagaaagttagttactaaatttttc
15 aacgttaatgaaaagaaaagaaccgtcataaaattaagagaagtgaacactatccaca
ttgattcaaaatttaataagttatgtcgtatggtttattgtccttacgtcaatactttca
cgtttcgggtattagtgtatcagcaattttagcaggagctggagttgttggtgttgccgtt
ggtttcggagcacaaacaattgtaaaagacattattactggtttctttatcatatttgaa
ggacagtttggatgtgagtgattatgttcaaattaatgcatctggggtaacaattgctgaa
20 ggtacggttaaaacgattgggttaagatcaacgcgtatacaatcagatactggagaaatt
tatacattacctaattggtatgattagtgaatagtttaattattctgctacagatgtttca
cctattgtgatgataccgattttctcaaagtgaattataaagtgatagaagagaaatta
ttaacattttttacctacattaaagaataaatatgacataattgtatccgcaccagattta
cttggttttagatagtgttgatggcaatgaaatggtgattaaacttttagcacatgtaaag
25 cctggaatgcattttccaggacaacggttacttcgtaaaagaggtcatacaatacttttagt
gaagaaggcattcatattccgaaaccaacacttgtaaaacttgataaagaattgaataaa
aaagaatag

Sequence 586

30 MTYCSLSIKIYITLKILIVLTIGGHVIMSQFKDTLYKLFEPMMKIEFYQNLLVNLIIILA
YILMGMIIVIAISRKLVTKFFNVNEKKNRHKIKRSETLSTLIQNLSYVWFIVLTSILS
RFGISVSAILAGAVGVAVGFGAQITVKDIIITGFFIIFEGQFDVSDYVQINASGVITAE
GTVKITIGLRSTRIQSDTGEIYITLPNGMISEIVNYSATDVSPIVMIPISPENYKIVIEKL
LTFLPTLKNKYDIFVSAPDLLGLDSVDGNEMVIKLLAHVKPGMHFPGQRLLRKEVIQYFS
35 EEGIHIPKPTLVKLDKELNKKE*

Sequence 587

Contig_0493_pos_9804_8986,
putative peptide of unknown function
40 atgaagatgacaaaacgcttaattatttcgactttgatgaaacttactacaaacatcat
acgcataagcgaatatgccttatttaagagaaatggaaggtttattacagaatataact
actaaaaacaatgtcattacggctattttaacaggaagtactatagaaagcgtaactcaa
aaaatgagtaacggttggtatgtcatataaacctcaacatattttttcagatttaagttct
aaaatgtttacatggaataactgtgaatatattgaatctgatgaatataaaaaacgaagt
45 ttgacagaacggtttcttattggaagatatattagatatataaaacatgtttcttctaaa
cataaagtagcgtttataccacaaagaacttttcgagacaatgaaacattgtacaatttc
tatctctattcttcgggtgacacgcatttagataaaacaattttagaagacctcactcag
tattctaaagataagggactatacgtatgacatttaacggttgtaatccttttagcaggtgat
cctgaaaatgcttatgatattaattttactccaagaaatgcaggaaaattatatgccaca
50 aaatttttgatgaataaatatggtgttccaaaagaattgattattggctttggtgatagt
ggtaatgatgaagcggtttttaagttatttagatcacgcaatgattatgtctaacagtcaa
gatgaggaaatgaagcgtaaatttaaaaaatacaaaatatccttattacaaaggtatttat
acacatgtacgcgaatttatagaatctgataatgtttaa

55 Sequence 588

MKMTKRLLIFDFDETYKHHHQANMPYLREMEGLLQNITTKNNVITAILTGSTIESVLQ
KMSNVGMSYKQHI FSDLSSKMFTWNNCEYIESDEYKNEVLTERFLEDILDILKLVSSK
HKVAFIPQRTFRDNETLYNFYLYSSGDTHLDKTILEDLTQYSKIRDYTMTFNRCNPLAGD
PENAYDINFTRNAGKLYATKFLMNKYGVPKELIIGFGDSGNDEAFSLYLDHAMIMNSQ

DEEMKRKFKNTKYPYYKGIYTHVREFIESDNV*

Sequence 589

Contig_0493_pos_7713_6802,

5 putative peptide of unknown function

atgaaaaaacatgtgtagcaatcgagcaattacagtaacaacattacttgcaggatgc
gatttttgagatttggtaggacagcatcagactgataagcaatcagaaaatagtaacact
caaaccgagcaagcttcaaataataaaaaattcaaattctaataatggtcattcaagta
aacaataagctagagatagagtcgaggatttaactcaatcgcaaaaagtagcattagct
10 atcaatgatccctcagtttctcaatatgccgttaacgcaagtgaattaagaaatcattca
ttttatgcaaaactataacggtgggggccaaacgtaaaagtattcatacgtatcagttggaa
gcattaccaacaaaagtagaaggtgcacctagtgatgaaattctatactgcaaaacca
tctaagggatcatttgtgacgcttatcggtatttggaatgaaaaggatttaattgcaggc
acacaaagttcaggaacatatcaacaatatgcgcattcggaagcagcaagagaattagat
15 ttacatgaactattaaataaatacggtaaaagttaaattataaagatatagcaaatcaa
attgcgtttacacaaagtcaatcatcaaatcaagtgaatctaacacgtagatgaaggg
acgtctaacagtgatagtgatacatctaattgatgataaagttacacgtagtaattgtgata
gataaagttgaagcgtatgaaggtcaccattagatactgatacatatacatttaaagaa
cccgaacagaacgaagacggagattggggcttttctatttttagataaggaaggcaacctc
20 gaaggttcttatattgtgacatctgatggtgaagttacaaaatacgaatgaaaacggggaa
gaaatagagtaa

Sequence 590

25 MKKHVLAIGAITVTTLLAGCDFGLVGHQTDKQSENSNTQTEQASNNKNSNSNNGHSSN
NNKSRDRVEDLTQSQKVALAINDPSVSQYAVNASELRNHSFYANYNGGGQRKSIHTYQLE
ALPTKVEGAPSDMKFYTAKPSKGSFVTLIGIGNEKVLIAGTQSSGTYQQYAHSEAARELD
LHELLNKYKSSNYKDIAHQIAFTQSQSSNSSESNTSDEGTSNSDSDTSNDDKVTRSNVI
DKVEAYEGHPLDTDYTFKEPEQNEGDWGFSLDKEGNLEGSYIVTSDGEVTKYDENG
EIE*

30

Sequence 591

Contig_0494_pos_7785_8111,

putative peptide of unknown function

gtgctagacacatcttttagctttatcaagtgcttttgacatgtcttttgctttgtca
35 aacgtagtaccacatgtctcttgcctttaccaccaatgtcttttactttattaacatca
ttttttacaccagagctaataacatctaataaccgtctttcttttttagtacctttacta
aatttaggtagtttttcttcttagtgatcatatcctgaattagataaaatagcatgtgtt
tgcgcgccactatatacacttgagccttttaggtaagaatggtgtgtatctttgttaggt
40 gttagtgccattcttccgttaggataa

40

Sequence 592

VLDTSALSSAFVPMFALSNNVPPMSLALPPMSFTLLTSFFTPELITSNIPSFVLVPLL
NLGSFFFLVSYPELDKIAVCAPLYTLEPLGKNVVVSLLGVSAILPLG*

45 Sequence 593

Contig_0494_pos_8613_9062,

putative peptide of unknown function

atgaaactgccaaactgctttaaatatgccaaattgtacctttttgagtgcatccatgta
tttttaactccagaccataatgcttttgctttattcactactgattttttatagccgctc
50 cagatttttaactgcggcattctttacagcattaaatattacaacaataccttttttaaat
gcgttaaatacagatagaacacctttgcgcaaaagctcgaacgattcctaacacaccattt
tttaatgcagtcacaaactttaatagagaaactcttaatagcattaaatatcgtaactact
atgcgcttaataaggttgatattaaattttacttgcgcaacatatgctttaataattgct
ataacaccgttctttaaggcggccagattttaatagcagcatttttcattgccattccat
55 aaagctgataagacattttttaatgcttga

Sequence 594

MKLPTALNMPIVPFLSAFHVFLTPDHNAFALFTTDFFIQVILTAFFFTALNITTIPFFN
ALNTDRTPLRKARTIPNTPFFNAVQTLIEKLLIALNIVTMMRLIRLILNFTCATYALIIA

ITPFFKAVQILIAAFFMPFHKADKTEFFNA*

Sequence 595

Contig_0494_pos_15336_15983,

5 putative peptide of unknown function

gtgtctacttcccaattgattgtttcgaattccggacgagctaactcaggggtttttctct
aattcagcaacagtggtgaatttagcggttagcacgttttaagatcgggtatttccactt
gcagttgatactgaagttttttgtaccaattctgataagtccttgactgtcttaacttct
10 ttttcaggaatatatttaatatcctctgggatagttacgccaacgtcatcagatttaaca
ttgtcacgttttagcccttttgatttcatgtactgttcaaagtctagaatttcttcggtt
gtctctggattttggtttaatttagccatagaacgtttcgctccttctttttgtctttt
tctttttttaattcttctctgttggttcttctactttttcaatagtaggtgtttctggt
gtttcttcaggtttgtcatctggttttggtgcatcatcaggtttttctcatctgaagtt
ccttctggttcatcatcagaaggtttgttctctgattcttctccagaattaccatctttg
15 ttatcttcaactctgcaccttcatctttaggtggttcatcttgtttaggtgctgacgct
tcaatttcttttgaaagctgttcgagttcttcgtactctttcttttga

Sequence 596

VSTSQLIVSNSGRANSNGFFSNSATVLNLALARFKIGYFPLAVDTEVFCTNSDKSWTVLTS
20 FSGIYLISSGIVTPTSSDLTSLRLAPFDFMYCSNARISSFVSGFWFNLAIERFAPSFLSF
SFFNSSSVGSSTFSIVGVSGVSSGLSSGFGASSGFSSEVPSSGSSSEGLFSDSSPELPSL
LSSTSAPSSLGGSSCLGADASISFESCSSSSYSFF*

Sequence 597

25 Contig_0494_pos_16481_16140,

putative peptide of unknown function

gtggatgaaaaagggtatactttaaatgccacttacctaatacatcatacgcaagagat
atztatgagaatattaaagcaggcaacgttaatcagtgacgtttcttttacacattgcca
cctaatactgactcaacggctcgtacgtggcaaaacatagataatgagtacgttcaaaccata
30 aataaaatcgatgaattgattgaggttagtattgttacagtgccagcctacaaagataca
tcggttgaaagtcggtcaacgtgcgaaagacttaaaagaaattcaaacagttggaacaaatg
aagatagcattggatttagaaagcctacgttttgaaacgtaa

Sequence 598

35 VDEKGLYFKCHLPNTSYARDIYENIKAGNVNQCSFFYTLPPNDSTARTWQNIDNEYVQTI
NKIDELIEVSIVTVPAYKOTSVEVGQRAKDLKKFKQLEQMKIALDLESLRFET*

Sequence 599

Contig_0494_pos_13618_13271,

40 putative peptide of unknown function

atgaatcatgttcaaaagaacaatattaaattctttgattatccaaacgcacaagaaatt
agagatgtagtgattgtcatagatccattaggccaagacaccctataacttacggagat
gatttccctatagcctttgaagatttgatcaaatagacgtgtttgtgaagcaacaaaaa
aacatcaatggaagattaacagctaaaaaaataacttttgagatagctaagggttttacga
45 acgataaatgtatatgacactggcggagctataaaagcctgaatatataaaagattttaat
atttacagacaaattaaaagatttgaaagtaagcaatcactcgtataa

Sequence 600

50 MNHVQKNNIKFFDYPNAQEIRDVVIVIDPLGQDTPITYGDDFPIAFEDLYQIDVFKQQK
NINGRLTAKKITFEIAKVLRTINVYDTGGAIKPEYIKDFNIYRQIKRFEVKQSLV*

Sequence 601

Contig_0494_pos_6051_5587,

putative peptide of unknown function

55 gtgacttattctgtctacaaaggatatgcagaatcattaaaagatacttctgaatttagt
tggactgatgaaagttggcaatttgacaaggtgttataggaagtgatgaagttaaata
aaacacaatattcgttactttaaaatatttaacggttctaaagatactattaacccttta
ttaagacacaaaattaaatattaattgcacacttacagcaccttatggatttgaaatcggt
aatctaaccacaaatgatataatttgaaatataaaaaaccgctcaaaaagcgtaatacggtt

tctattataggagtgcatccttatattaataataaaaagagtttggtaaaagacacaaattat
gattttattacttttagcgccgggttggaatgaaattttaattagaggtcacaaatatcc
aatagtcctaaaaacagaatttatatttaattacatctataggtag

5 Sequence 602

VTYSVYKGYAESLKDTSEFSWTDESQWQFEQGVIGSDEVKYKHNIRYFKIFNGSKDTINPL
LRHKLNINCLTAPYGFEIVNLTTNDIFEYKKPLKKRNTVSIIGVHPYINNKRVGKDTNY
DFITLAPGWNEILIRGHNISNSPKTEFIFNYIYR*

10 Sequence 603

Contig_0494_pos_5586_4018,

putative peptide of unknown function

gtgagaatattggaaaatctaataatttatgaatagagaaggacattttcggaaattggt
aatgactttgattttggttcctttaaatatgaatatgaacaaaataatgagcgatccata
15 tctctcactgcttataaaaactaatgttaacgcgatataatttgatagtttgattaatgaa
aattatttagtttggaagggccagaaatatgtcattaaatcgactgagcttaagtatgaa
gaaggtgtaataacttaatgaaattgaggctaaagcatatttctatggaatttcaaaatcat
tatatacctaagatttttagatgatgagtcactgaatgatgaagatgagactgaagcaaaa
atttccatgaaagttaaagagtaccttgattttgcattcaaaaaataataaacttaatttc
20 gattataagttacatggaaaatttaattgagagtaaatatattgaacagttaggagataaaa
aatggtttagaacatcttattgaaggtgctgagcattttggctatatatttttgctgat
aataaaaactttccatatctatacacctgataatttttataaaaaaatcagatgaaatatta
gtttataaataataaataagttcgggtttcggctaaaacaatcacaaactgaattacgcacc
tacattcaaggatatggaaaagaaagtcgaaatccgaaacgaaaaactataaacctata
25 aaacctaagcattttctcatactctggaaattttaataaagaagggaacttggtctactgaa
catataggagattcggttttataagacatttgattgtaagtgggggaatgaaaccttaact
tggaatctaaaaaaaggacctaaggtggaattatcgaagtatttattgatgataagtcg
aaagggaacttttgattgttacagcgctcatgcttcgacgcaaaaagtgatttttagctaaa
ggattatcaaaaggtaaacattcttttagaggagtttttaaatcgaaaaaacctggtatt
30 gattataagaagtctaataccagtcagtgatgttggtacgagtaaaagttagtggttttaaat
ctaactcgagttcttaaaaggtaaagatatattcatgtatatgctgaatataagtcctcca
tattataagcaatatggtaaatcagaagccctacaatatatgatgataatattacaagt
caatcagagttaaagaagaaattaaaagaaacacttgatgacataccaacaatcgaagta
gcaacgaattatttaggattagaaagtattcatgaaaataataactattcgattttatacac
35 aaacctatcggatttaataactgatttaaaagtgtgcaaacctactgaatatcaccctctt
gtttcgcagcctattgaagtggaattcagtaatgcacagaaagatatataaaaaatgcaa
tcacagttcaatcgtaggttaagaaaggttaataatcttatgaaaaaggattcaaaact
agtgactattctttaaatgtgttagaggaatataacgaaacagtaggaagtgtattgatt
gatgagtaa

40

Sequence 604

VRILENLI FMNREGTFSEIVNDFDFGSFKYEYEQNNERSISLTAYKTNVNADIFDSLNE
NYLVWKGQKYVIKSTELKYE EGVILNEIEAKHISMEFQNHYIPKDLDES LNDEDETEAK
ISMVKEYLDFAFKNNKLNFDYKLHGKFNISKYIEQLGDKNGLEHLIEGAEHFGYIFFAD
45 NKTFHIYTPDNFYKKSDEILVYKYNSSVSAKTITTELRTYIQGYGKKKSKSETKNYKPI
KPKDFSYSNGFNKEGTWSTEHI GDSFYKTFDCKWGNETLTWNLKKGPKGGIIEVFIDDKS
KGTDFDCYSAHASTQKVILAKGLSKGKHSFRGVFKSKKPGIDYKKNPVMYVGTSKSSVLN
LTAVLKGKDIYHVYAEYKSPYKQYKSEAPTYYDDNITSQSELKKKLKETLDDIPTIEV
ATNYLGLESIHENNTIRFIHKPIGFNTDLKVVKLTEYHPLVSQPIEVEFSNAQKDIKMQ
50 SQFNRRRLRKVNLMKKGFKTSDYSLNVLEEYNETVGSVLIDE*

Sequence 605

Contig_0494_pos_3836_1974,

putative peptide of unknown function

atgttattaaacttttagactttcctattcaaataggacacacatttagaaccaagatgata
55 aataatttttagaacaatacttaattattataatgaattagatcatcagcatcgcgacac
acagaaaactaagcatcatgcacatcaagccatgcaggttgattatagaaatacaaacggtt
tctgcatttttagattatcttaacggtaatatgaatgggcttgttttaggagcaaatgga
gacggtagatctgaaacaaaacaagccagagtatcaatagatggtaccgtacatcccttg

ttgcaagaaaggctgcttcatgacttttttaggaattaaacagaaaattagataaaagaaata
 cattctaattggtgcagttgattttatttggaaatcctccatatataccaggaaatagattg
 ggagaaaatgggacaccaaataattgggaaccagaagcccatattgaagcgttttttaaac
 cctttagttgataatcaatacgttacaaaagaagttaggagaagatacatcaggaaaa
 5 tataatgtgtacaaaatttacgtttgaaccacaaaattacaataaaacgttacttattact
 tcatgtatacacggtaaatgaaactactggattttttgatatgtgccatatactcaatcta
 ttgggtcaatcaatgggaaaagtatcctcaattaacttacttaagaaaaaatgtacgttta
 atttatgttccataggttaacccgtggggattcgcaaatcaagaaagagagaatgtgaac
 aatgtagattttaaacagaaaattttgattataactggaaggcaggtaaaggacagatcct
 10 gataaatctaacttcaaaggtaaaagtccttttctgaaaaagaatcacaaaatatgcgt
 agcttagttcaaagtatagataatttaactgctcacttagatttgcatgatattatttca
 gtaaatgaatgattactgtttattttatccgcttgggccaatcaaaaaataataaatg
 actcatcttattaacaatttaaaaagtaacggagacctcgttggttggggtccagtaca
 ttatcatcttttagtaattgggttaggaatccgaaataaaacaacgtcatatctttcagaa
 15 ataaatgaaaaacgtgtcgggtgaaaagaaaagtcgccgaagaatgagacgttcagtagc
 tgggttaggtaatttttagaatggcacaatttgaatcttatcaaaatggtcaaaca
 tcatttagatcctttcattaaagtgtatggtatgatgatagatttaacaataaaacatct
 gaagtcatcaccctacgtgcagaaaggaatgaatggcaacgtataatgatgagtcagcag
 cgtttcaaagtcttttagcaaatggattttagtagctctatggatatgtgactataaacgtt
 20 gatagagatgtcacagtggtggttaatacctaataattgttcagaattatcatccattcttt
 ggatttaataaaaagtagaaaacgtaattttttcaattgaacatagactcaacaaagga
 aatacaactttccctattttacgctgctgctggagttcaaatgtcgacgattactgaacca
 ggtacaaaacgtactgatacagtaatgccggtacttagatgttaagaaaaaagggtgctggt
 attgtaacaatcaacaaaattaaattatttgcgaagttcactcctacgcattctgcta
 25 tccattcagatattaaaatctggagaatcggtaatcttaagaagatacgttcacacaa
 attaccctaatactatatatgatgatgatttaagaaatgttataaatggggaggaaaaa
 taa

Sequence 606

30 MLLTLDFPIQIGHTFRTKMINNFRTILNYYNELDHQHRAHTETKHHAHQAMQVDYRNTNV
 SAFLDYLNNGINGLVLGANGDGAETKQARVSIDGTVHPLLQERLLHDFLGINRKLOKEI
 HSNQAVDFIWNPPYIPGNRLGENTPNNWEPEAHIEAFLNPLVDNQYVTKEVIGEDTSGK
 YNVYKFTFEPQYNKTLITSCIHGNETTGFFDMCHILNLLVNQWEKYPQLTYLRKNVRL
 IYVPMVNPWGFANQERENVNNDLNRFNDYNWKAGKGTDPDKSNFKGKSPFSEKESQNM
 35 SLVQSIDNLTALHDLHDIISVNNDYCLFYPRWANQKNNNMTHLINNLKSNGLVWVGSST
 LSSFSNWVGIRNKTTSYLSEINEKRVGEKKSPEEMRRSVRWVGNVIFRMAQFESYQNGQT
 SLDPFIKVMVYDDRFNNKTSEVITLRAERNEWQRIMMSQQRFKVLANGFVELYGYVTINV
 DRDVTVGINPNIVQNYHPFFGFNKSRRNLFSEHRLNKGNTTFPIYAAAGVQMSITIEP
 GTRKTDVTMPVLDVKKKGAGIVTIKQIKLFAKFTPTHSANSIQILKSGEYGNLKEDTFTQ
 40 IYPNTIYDDDLRNVINGEEK*

Sequence 607

Contig_0494_pos_1581_760,

putative peptide of unknown function

45 gtgggtgacgcaataataaaaaaataatggtgcaactaaaaattaaatatatccgtatg
 tttgatgaattaaaaagacaaattaatgcacgagccactgaaatacaagaacaattagat
 aatttagaagattacgttggttaaagtgaagatgcaagtgtgaaggaattacaaagatt
 cagattgaacacaaaaaaggattggacaaacttaataacacagcgtagtaaaagttaaaa
 gacgtcgaggaatctcttaacgcggctaaaaatacaattcaaaatctttatgaagaatat
 50 gacaacgaaattgacacaaaaaggaagtcaatatttaaaagatttaagaatcgaagtagg
 aatattgaaaatatattaagtcaagagggtacgtcacaaattgatgaacatcgtaaaagc
 attactgaaatacaagaaaagtacctgaatcttcagactggattgaatatgatttgatt
 aatggagctataaaaaatagcattataaagctgaaggacaaaatgggttttaattgcgct
 tataaaaatcattcaacatcaagactataaggaagtgtgtaagaattaacgctgacaa
 55 tttaaaagtgaactgttatagcgaagttaccgagtgaaactaattacaagtacgcaaac
 gcgttcctaagatcggtgcctgttaaagcttgggtgctcaattaactattgaaccta
 ggagatgttaaagtttatatttctcagagcgatcagtggtcagtaagtcgtgaagcttat
 atttacggagaaattagaatgatagataaaggagggtgaataa

Sequence 608

VGDAlINKINGATKIKYIRMFDELKRQINARATEIQEQLDNLEDYVVKVKDASDEGITKI
 QIETKKGLDKLNQQRSLKDVESLNAAKNTIQNLYEYDNEIDTKGSQYLKDLRIEVR
 NIENILSQEGYVTIDEHRKSITEIQEKLPESSDWIEYDLINGAIKNRHYKAEGQNGFNCA
 5 YKIIQHODYKEVMLRINADNFKSGTVIAKLPSELITSTQTAFLRSVPVKACGAQLTIEPN
 GDVKVYISQSDQWSVSREAYIYGEIR MIDKGG*

Sequence 609

Contig_0494_pos_758_132,
 10 is similar to (with p-value 3.0e-23)
 >pir:pir|S41182|S41182 hypothetical protein 37.1 - phage SPP
 1 >pir:pir|S43808|S43808 hypothetical protein 38 - phage SPP
 1 >gp:gp|X67865|BSSPP1_10 B.subtilis phage SPP1 DNA sequence
 coding for products required for replication initiation. NI
 15 D: g472886.
 gtgatggatacttataaatctatgactgaacttgtgaggaatgaaaagattggatgatt
 gagacacaagatagaaatagtaaatcacttataactgctatacacggaggcggtatagag
 tgtggcacttctgaattagcgttattggttgacagaattatcgaaatgcaaaactatcttact
 tttaaagggtttaaaaccgaaaaacaatagaactctacacgtcacttcaacaaattatgat
 20 aaccccaattttattatattggaatcaattttatgaatgtaacgatagccgtacatggttat
 tcgagcaatcaagcaaatagttatattggtggattggatgaaagacttatatctcttatt
 actcacaattttaaaagtttcagggttttaattgtggaagctgctcctgacagaattgcgggc
 agagaaattaataatataaccaacaaaaatgcctatggcatgggtgtacagattgaaata
 tcgactcaacaaagaaaaagaatttttttagtcgaaacgatttttagtaaaaagaatagagaa
 25 aatacacataattggacagaagatatgtattattatgctaattgctatttgtgctgcactt
 aatgatagaaagtgggtagaaacatga

Sequence 610

VMDTYKSMTELVRNEKDWMIETQDRNSKSLITAIHGGGIECGTSELALLVAELSNANYFT
 30 FKGLKPKNNRTLHVSTSTNYDNPNNLYWNQFMNVTIAVHGYSSNQANSYIGGLDERLISLI
 THNLKVS GFNVEAAPDRIAGREINNITKNAYGMGVQIEISTQQRKEFFSRNDFS SKNRE
 NTHNWTE DMYYYANAICAALNDRKWWET*

Sequence 611

Contig_0495_pos_3147_2617,
 35 putative peptide of unknown function
 atggtgctcgatcaatttccaccttggtttgattgtaacgtccaaaatataaattacatc
 ttatatgtgagaaaacaattactgatattccgatgagcattgaatttagacatcaatca
 tggtttgacaatcagtataaagaacaaactttatccttcttaacacaacatcaaatcatt
 40 catgcagtggttagatgaacctcaagttaaagaggggagcggttccttttagtaaataggatt
 actagtgaattgtctttgtacgttatcatggacgtaatacattatggttggactaaaaaa
 gatatgactgatcaagaatggcgagatgtaagatatattatgattatagcgatgatgag
 ttagctgacttggctcgtaaagtcgaaatacttaatacaaaaggctaagaaagtatatgta
 atttttaataataactctggcggtcatgcagctaataatgctaaaaagtatcaaaatatt
 45 ttagacattgattatgaagggttagcaccgcaacaattaaaactatttttaa

Sequence 612

MVLVQFPWFDCNVQININYILYVRKQLTDIPMSIEFRHQSWFDNQYKEQTL SFLTQHQII
 HAVVDEPQVKEGSVPLVNRITSEIAFVRYHGRNHYGWTKKDMTDQEWDRVRYLYDYSDE
 50 LADLARKVEILNQAKKVYVIFNNNSGGHAANNAKKYQNILDIDYEG LAPQQLKLF*

Sequence 613

Contig_0495_pos_2265_1777,
 putative peptide of unknown function
 55 gtgtttatgatattttgtgctgatattactgatgattcgctcacaaaatcaaaccttttaaa
 atttttgacaaacctaaatatgcgcgtacatatgttgatgctgaagggaacataaccgt
 tatagtgtaccacccctgtttgcttttataacaacgttattttattgggctattaacagga
 ctgtttggcataggtggaggtgcattgatgacccctcttatgctcatcgctcttttagattt
 ccaccacatggtgagtaggcacaagtatgatgatgattttcttttcaagtgtgatgagt

tcaatagggcacatctttcaaggacatgtggcttggggctattctatcattctcattatt
 tcaagtgttataggtgcacaaataggtgtgagggcgaatcgatctatgaaatccgacaca
 gttgtaatgttattgagaacagtaatgcttatcatgggtgtatatattaatcattaaatct
 tttatttaa

5

Sequence 614

VFMIFVSILLMIRHKIKPFKIFDKPKYARTYVDAEGKTYRYSVPPLFAFITTFLIGLLTG
 LFGIGGGALMTPLMLIVFRFPPHVAVGTSMMMIFFSSVMSSIGHIFQGHVAVGYSIILII
 SSVIGAQIGVRVNRSMKSDTVVMLLRTVMLIMGVYLIKSFI*

10

Sequence 615

Contig_0495_pos_1181_372,

putative peptide of unknown function

atgagccatgtcgggtatctttttgatgaaaagttatgccaaagagattccggaaatagat
 15 gttatcttttggtagtcatacgcacatcatcttttgaacatggagaaataaacaatggtgtt
 ttgatggcagctgccggaaaatatggctattatttaggtgaagttaattacgattgaa
 aatggaaaaatcggtgataaaatcgccaaaattcatcctattgaaacacttcccttagtc
 gagacacattttgaagaagaaggaagagcacttctaagtaaaccagtagttaatcatcat
 gtgaacttagtcaaaagaacagatgttggttacaagaacatcgatattactggctgaaagt
 20 gtatatgagttttcaagggctgattgtgcaatcgtaaatgctggacttatagttaatggc
 attgaagctgataaagtgcggaatatgatatacatcgcatgttaccatccaatcaat
 attgtaagagttcgattaaccggtaaacaattaaagcaagtgattcaaaaaagccaaaag
 caagaatatatgcacgaacatgcacaaggtcttggttttagaggggatataatttgaggga
 tatattttatataatctaggtcttattgagtcagaagaccgttattttataggcgatgaa
 25 gagattcaaaatgataaacaatatacggttaggtactgttgatatgtatacatttggaaga
 tatttccattgtcgaaggggttatctacagattatattatgcctgaatttttacgtgat
 atttttaagagaaattactaaaattataa

Sequence 616

30 MSHVGIFFFDEKLQEIPEIDVIFGSHTHHHFEHGEINNGVLMAAAGKYGYLGEVNITIE
 NGKIVDKIAKIHPIETLPLVETHFEEEGRALLSKPVVNHVNLVKRTDVVTRTSYLLAES
 VYEFSRADCAIVNAGLIVNGIEADKVTEYDIHRMLPHPINIVRVRLTGKQLKQVIQKSQK
 QEYMHEHAQGLFRGDIFFGGYILYNLGFIESEDYFIGDEEIQNDKQYTLGTVDMYTFGR
 YFPLLKGLSTDYIMPEFLRDIKELKLLK*

35

Sequence 617

Contig_0496_pos_0_1167,

is similar to (with p-value 0.0e+00)

>sp:sp|P39772|SYN_BACSU ASPARAGINYL-TRNA SYNTHETASE (EC 6.1.
 40 1.22) (ASPARAGINE--TRNA LIGASE) (ASNRS). >gp:gp|L47709|BACYP
 IA_24 Bacillus subtilis (clone YAC15-6B) ypiABF genes, qcrAB
 C genes, ypjABCDEFGH genes, birA gene, panBCD genes, dinG g
 ene, ypmB gene, aspB gene, asnS gene, dnaD gene, nth gene an
 d ypoC gene, complete cds's. NID: g1146223. >gp:gp|Z99115|BS
 45 UB0012_176 Bacillus subtilis complete genome (section 12 of
 21): from 2195541 to 2409220. NID: g2634478.
 atgaaaactacgattaaacaagcgaaaaaacatcttaaccaagaagtaacaattggtgct
 tgggttaactaataaaacgttcaagtggttaaaatagcgtttttacaattacgcgatggtaca
 ggatttatgcaaggagtagtagtaaaatctgaagtagatgaagaaacatttcaactagca
 50 aaagatataactcaagaatcatctttatacatcacaggaacgattacagaagataatcgt
 tctgatttaggtacgaaatgcaagttaaatcaatcgaaattgtacatgaagcacacgat
 taccctattacaccaaagaatcatggaacagaatttttaattggatcatcgtcacttatgg
 ttacgttcaaaaaaacaacatgctgtcatgaaaataagaaatgaaattatccgtgcaaca
 tatgagttttcaatgaaaatggcttcactaaaattgatccacctattttaacagcaagt
 55 gcaccagaggggaacaagtgaagttattccatacaaaaatatttcgatgaagatgcattctta
 tcacaaagtgggcagttgtatatggaagcagccgcaatggctcacggacgtgttttttca
 tttggcccaacttttcgtgcagaaaaatctaaaacacgcgcgtcatttaattgaattctgg
 atgattgaaccagaatggcctttacaaatcatgcagaagcttagaaatacaagaacag
 tatgtgtctcacattgttcaatctgttttaaatcattgccaaattagaactcaaagcttta

gatagagatacaactaaactagaaaaagttgctacacctttccctagaatttcttatgat
 gatgctatcgaaattcttgaaaaaagagggattcgatgatattgaatggggtgaagacttt
 ggtgcacctcatgaaacagccatcgctaatactatgatttaccagtatccattacaaat
 tatccaactaaaattaaaccattctatatgcaaccaaaccagacaatgaagatacagta
 5 ttagtgctgatttaattgcgcctgaaggttacgggtgaaattattggtggtccgaacgt
 attaattgatttagaattattagaacaacgcattaatgagcacgaattggatgaggaaagt
 tatagctattatttagatttacgtCTT

Sequence 618

10 MKTTIKQAKKHLNQEVITIGAWLTNKRSSGKIAFLQLRDGTGFMQGVVVKSEVDEETFQLA
 KDITQESSLYITGTITEDNRSDLGYEMQVKSIEIVHEAHDYPITPKNHGTEFLMDHRHLW
 LRSKKQHAVMKIRNEIIRATYEFFNENGFTKIDPPILTASAPEGTSELFHTKYFDEDAFL
 SQSGQLYMEAAAMAHGRVFSFGPTFRAEKSKTRRHLEFWMIEPEMAFTNHAESLEIQEQ
 YVSHIVQSVLNHCQLELKALDRDRTTKLEKVATPFPRI SYDDAIEFLKKEGFDDIEWGEDF
 15 GAPHETAIANHYDLVPFITNYPYPTKIKPFYMQPNPDNEDTVLCADLIAPEGYGEI IGGSER
 INDLELLEQRINEHELDEESYSYLDLRL

Sequence 619

Contig_0497_pos_6106_5558,
 20 is similar to (with p-value 4.0e-90)
 >sp:sp|P51183|PT1_STAAU PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOT
 RANSFERASE (EC 2.7.3.9) (PHOSPHOTRANSFERASE SYSTEM, ENZYME I
). >gp:gp|X93205|SAPTSHI_2 S.aureus ptsH and ptsI genes. NID
 : gl070384.
 25 atgttcccaatggttagcaacaattaaagaattccgtgacgctaaatcaatgcttcttgaa
 gagaaagaaaaatcttcttcgcgaaggttacgaagtttcagatgatattgaattaggtatt
 atggttgaaattccagctaccgcgcgcacttgctgatgtatttgctaaagaagtagatttc
 ttttagtataggaacgaatgacttaattcaatacacattagctgctgaccgtatgtctgaa
 cgagtttcatacttatataccatataatccttcaattttacgattagttaaacaagtt
 30 attgaagcttctcataaagaaggtaaatggactggtatgtgtggtgaaatggctggagat
 caaacagctgtgcctttattattaggttttaggttttagatgagttctcaatgagtgcgact
 tctatcctaaaagctagaagacaaatcaatggtttaagtaaaaaatgaaatggctgaactc
 gctaataagactggttgatgctcaacgcaagagggaagtcggttgatttagttaaccaatta
 gctaaataa
 35

Sequence 620

MFFPMVATIKERDAKSMLEEKENLLREGYEVSDDIELGIMVEIPATAALADVFAKEVDF
 FSIGTNDLIQYTLAADRMSESVSYLYQYPNPSILRLVKQVIEASHKEGKWTGMC GEMAGD
 QTA VPLLLGLGLDEFMSATSILKARRQINGLSKNEMAELANRAVE CSTQE EVVDLVNQL
 40 AK*

Sequence 621

Contig_0497_pos_2347_1763,
 is similar to (with p-value 3.0e-69)
 45 >sp:sp|P39760|YKQB_BACSU HYPOTHETICAL 24.3 KD PROTEIN IN KIN
 C-ADEC INTERGENIC REGION (ORF4). >gp:gp|AF012285|AF012285_27
 Bacillus subtilis mobA-nprE gene region. NID: g3282109. >gp
 :gp|D37799|BACAMOKOOO_6 Bacillus subtilis genes for ampS, mr
 eBH, orf1, kinC, orf3, orf4 and orf5. NID: g520838. >gp:gp|Z
 50 99111|BSUB0008_123 Bacillus subtilis complete genome (sectio
 n 8 of 21): from 1394791 to 1603020. NID: g2633699.
 atggatgttatggctattgatagagacgaaaatcgtgttaacgaatatagtgatatagca
 acacatgcagttgttgctgatacaactgatgagcagtaataagagtttaggaatacgt
 aatttcgatcatgttattgtcgctattgggtgagaatatacaatctagtacactaacgacg
 55 ttaattcttaaagaattaggtgttaaaaaggttactgctaaagcccaaatgattatcat
 gctaagatttttaataaaaataggtgccgatactgttgctgacccctgaaagagatatggga
 agacgtattgtcataatgttgctagtgtgctgtccttgactacttggaacttgctgat
 gagcattcaatcggttgaaatctacagaaaaatggcaggacaaacaattattgaa
 ttagatattcgagctcaatatggtattaacattatagcaattaaaagagctaaagaattt

atagtctctccagaccctaacaatcaatattgaaataggggacattttaattatgattggt
catgataatgacttaggtcgctttgaaaaaatataagcaagtaa

Sequence 622

- 5 MDVMAIDRDNRVNEYSDIATHAVVADTTDEAVMKSLGIRNFDHVIVAIGENIQSSTLTT
LILKELGVKKVTAKAQNDYHAKILNKIGADTVVHPERDMGRRIAHNVASASVLDYLELAD
EHSIVELKSTKEMAGQTIIELDIRAQYGINIIAIIKRAKEFIVSPDPNINIEIGDILIMIG
HDNDLGRFEKNISK*

10 Sequence 623

- Contig_0498_pos_2017_3027,
is similar to (with p-value 8.0e-80)
>sp:sp|P41006|PYRP_BACCL URACIL PERMEASE (URACIL TRANSPORTER
) . >pir:pir|S38893|S38893 uracil transport protein - Bacillu
s caldolyticus >gp:gp|X76083|BCPYRQP_2 B.caldolyticus (DSM40
5) pyrR, pyrP and pyrB (partial) genes. NID: g431229.
atgctggttgacattttatgagtggttaataatgtacgtgattataggtattttcattaaa
ttgagtggaacacattggttaatgcacttgttaccaccagtagttgtcgaccagtaata
atggctaatgggttaagtttagctcctacagcagtaaacatggccatgttcgaaaattct
20 gctgaaatgaaagggtataacttaagttacttaattgttgctttgattacattagcagta
accatcatcgtccaaggattcttcaaaggatttttatcactaataacctgtacttataggt
attatagtggtgatataattgtatccattttcatgggcatagttaaatttgctccaatagca
caagcgaaatggatagattttcctcatatttatctaccatttaaagattacacaccatct
tttcatttaggactcattctcgtgatgatacccggtggtgtttgtgacggtaagtgaacat
25 attggtcatcaaattgtaattaataaaatagtaggacgcaattttcttgaaaatccagggt
ttagataaatcaatcattggtgatggtgtttcaactatgtttgcaagtatgataggagggt
cctcctagtacaacttatggtgaaaatataggtgtactagcgatcaccaaaatatatagt
atttacgttatttggtggtgcggcagttatagctatcattcttgcatttattggttaagttc
actgctttaatatcttcaataacctacgccagtgatgggtggtgtctcaattttattattc
30 ggtattatagcagctagtggtttaagaatgcttggtgaaagtcaagtagatttcgcaagc
aatcgcaacttggttatagcatcagttgtgcttggtgctgggattggtaattcttcttacc
aatttaaaaggcataggtatcaatttacaattgaaggaaatggcattatcagcactttca
ggaataatattaaatttaattttgccaaaagataaaaaccaaataaattaa

35 Sequence 624

- MLVALFMSGLMYVIIGIFIKLSGTHWLMHLLPPVVVGPMVIMVIGLSLAPTAVNMAMFENS
AEMKGYNLSYLIVALLITLAVTIIIVQGFKGLSLIPVLIGIIVGYIVSIFMGIVKFAPIA
QAKWIDFPHIYLPFKDYTPSFHLGLILVMIPVVFVTVSEHIGHQMVINKIVGRNFFENPG
LDKSIIGDGVSTMFASMIIGPPSTTYGENIGVLAITKIYSIYVIGGAIVIAIILAFIGKF
40 TALISSIPTPVMGGVSILLFGIIAASGLRMLVESQVDFASNRNLVIASVVLVVGIGNLLI
NLKGIGINLQIEGMALSALSGIILNLILPKDKNQIN*

Sequence 625

- Contig_0498_pos_3053_3934,
is similar to (with p-value 2.0e-90)
>sp:sp|P05654|PYRB_BACSU ASPARTATE CARBAMOYLTRANSFERASE (EC
2.1.3.2) (ASPARTATE TRANSCARBAMYLASE) (ATCASE). >pir:pir|A25
015|OWBSAC aspartate carbamoyltransferase (EC 2.1.3.2) catal
ytic chain - Bacillus subtilis >gp:gp|M13128|BACPYRB_1 B.sub
tilis pyrB gene encoding aspartate transcarbamoylase, comple
te cds. NID: g143383. >gp:gp|M59757|BACPYROP_3 Bacillus subt
ilis pyrimidine biosynthetic (pyr) gene cluster (pyrR, pyrP,
pyrB, pyrC, pyrAA, pyrAB, pyrD, pyrF and pyrE) genes, compl
ete cds. NID: g387576. >gp:gp|Z99112|BSUB0009_20 Bacillus su
55 btilis complete genome (section 9 of 21): from 1598421 to 18
07200. NID: g2633902.
atggaacacttattatcaatggagcatttatctaattcagaaatttatgatttaattact
atcgcttgccaattcaaatctggtgagcgaccattacctaatttaacggtcaatacgtat
tcaaaccttattcttcgaaaattcaacgcgaacaaagtgtagctttgagatggcagaacaa

aaattaggattaaaaacttattaatTTTgaaacaagtacatcatctgtaaaaaagggtgag
 tcactttatgacacatgtaaaacacttgaaagtataggtggtgatttacttgtcatatgct
 cactcccaaaattcttattacgaagaactggatcaattaaatattccaattgctaattgca
 ggtgatggaagtggacaacatcctactcagagtttattagacataatgacaatatatgaa
 5 gaatatggttcggtttgaagggtttgaatattctaataatgtggggacattaaaaattctcgt
 gtcgcaagaagtaattatcatagtttaacatcattaggtgccaacgtaattgttctcaagt
 ccaaaagaattgggtagataatacattagaggcgcttatgttgaaattgatgaagtcatt
 gataaagtagatattgttatgttgcttagagttcaacatgaaagacatggaatttcaggt
 gaagctaactttgctgctgaagaatatcatcaacaatttggttttaacacaggctagatat
 10 gataaattaaaagaggaagccattgtaatgcatccagctcctgtaaatagagggtgttgaa
 attaaaagcgagctagttgaagcacctaagtcctcgaatatttaagcagatggaaaatgga
 atgtatttaagaatggcagtaataagtgcgcttttacaatag

Sequence 626

15 MEHLLSMEHLSNSEIYDLITIAQCFKSGERPLPQFNGQYVSNLFFENSTRTKCSFEMAEQ
 KLGLKLINFETSTSSVKKGESLYDTCKTLESIGVDLLVIRHSQNSYEEELDQLNPIANA
 GDGSGQHPTQSLDDIMTIYEEYGSFEGNLILICGDIKNSRVARSNYHSLTSLGANVMFSS
 PKEWVDNTLEAPYVEIDEVIDKVDIVMLLRVQHERHGISGEANFAEEYHQFGLTQARY
 20 DKLKEEAI VMHPAPVNRGVEIKSELVEAPKSRI FKQMGMYLRMAVISALLQ*

Sequence 627

Contig_0498_pos_3952_0,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P46538|PYRC_BACCL DIHYDROOROTASE (EC 3.5.2.3) (DHOASE
 25). >pir:pir|S34319|S34319 dihydroorotase (EC 3.5.2.3) - Baci
 llus caldolyticus >gp:gp|X73308|BCPYR_2 B.caldolyticus pyrim
 idine biosynthesis genes. NID: g312439.
 atgaaattaattaaaaacggaaaaatcttaaaaaacgggtatcctaaagacacagaaatt
 ttaatcgacggtaaacgtattaaacaaattagtagtaaaattaatgcttcatcttcaaat
 30 attgaagtatttgatgcaaaaggaaatttaattgctcccggtttttagatgttcatgtg
 cacctacgtgaaccagggtggtgaacataaagaacaattgaaagtggtaaaaaagccgct
 gcaagagggtggttttactacagtatgtcctatgcctaataacaagacctgtaccagataca
 gttgaacatggttagagaattaaagacaacgaatttctgaaacagcacaagttagggtgtg
 ccttatgctgctattactaagagacaagcaggtactgaacttgttgattttgaaaaatta
 35 gcactagaagggtgtgtttgcatttactgacgatggtgtgggagttcaaacagcaagtatg
 atgtatgctgctatgaagcaagctgcaaaagttaaaaaaccgattgtcgacactgtgaa
 gataatagcttaattctatggtggtgcaatgcataaaggtaaacgtagtgaagaattaggc
 atacctggtattocaaatattgctgaatctgtacaaattgctagagatgtattattggct
 gaagcaactgggttttactactatcatgtgtcatgtttcaactaaggaaagtgttcgagta
 40 atcagagacgcttaaaaaagctgggtatccatgtaacagcagaagttacaccacatcattta
 ttattaactgaaaatgatgttcctggcgatgattcaaactacaaaatgaatccaccatta
 agaagtaatgaagatagagaagcacttttagaaggcttattagatggaacaattgatgtg
 attgcaacggatcatgcacctcacgctaaagaagaaaaagcacacactatgacaaaaagca
 cctttcggcatcgtaggttagtgaaacagcattcccattactttatacacactttgtaaga
 45 cgaggttaattggctcactgcaacaattagttgattatttcactattaaaccagctactatt
 ttcaacttaattatgaaaattacacaaagat

Sequence 628

50 MKLIKNGKILKNGILKDTIELIDGKRIKQISSKINASSSNIEVIDAKGNLIAPGFVDVHV
 HLREPGGEHKETIESGTKAAARGGFTTVCPMPNTRPVPDTEHVRELQRRISETAQVRVL
 PYAAITKRQAGTELVDFEKLALEGVFAFTDDGVGVQTASMMYAMKQAAKVKKPIVAHCE
 DNSLIYGGAMHKGRSEELGIPGIPNIAESVQIAROVLLAEATGCHYHVCHVSTKESVRV
 IRDAKKAGIHVTAEVTPHHLLLTENDVPGDDSNYKMNPPLRSNEDREALLEGLDGTIDC
 IATDHAPHAKEEKAQPMTKAPFGIVGSETAFPLLYTHFVRRGNWSLQQLVDYFTIKPATI
 55 FNLNYGKLHKD

Sequence 629

Contig_0499_pos_4575_5165,
 is similar to (with p-value 3.0e-47)

>sp:sp|P42954|TAGH_BACSU TEICHOIC ACID TRANSLOCATION ATP-BINDING PROTEIN TAGH. >gp:gp|U13832|BSU13832_2 Bacillus subtilis 168 highly hydrophobic integral membrane protein (tagG) gene and ATP-binding protein (tagH) gene, complete cds. NID: g 755151. >gp:gp|Z99122|BSUB0019_67 Bacillus subtilis complete genome (section 19 of 21): from 3597091 to 3809700. NID: g2 636029.

atgattgggtggtctatatttcaccaagttccgggtgaaataacgagacatggtgatgtgagt
gtcatcgctattaatgcaggactaaatggacaattgacaggtgtagaaaatattgaattt
10 aaaatgctctgcatgggcttttaaaggaaagaaattaaaaaattaatgccggaattata
gaatttagtgaaactcggcggaatttatttatcaacctgttaaaaaatattcaagtggatg
cgtgcaaaacttggattttcaattaatattactgttaacctgacatattagttattgac
gaagcattatcagtaggcgatcaaacatttactcaaaaatgtttagataaaatttatgaa
tttaaagcggctaaaaaaacaatattttttgttagtcataatattagacaagtgcgtgaa
15 tttgtacaaaaatcgcttgattgagggcggtaaactaaaagaattcggcgaaacttgaa
gaagtattacctgattatgaggcggtttcttaaaacttttaagaaaaaatctaaagcagaa
caaaaggaatttagaataaattagatgagtcacgttttgtcgtaaaataa

Sequence 630

20 MIGGSISPSSGEITRHGDVSVIAINAGLNGQLTGVENIEFKMLCMGFKRKEIKKLMPEII
EFSELGEFIYQPVKKYSSGMRAKLGSINITVNPDIILVIDEALSVGDQFTTQKCLDKIYE
FKAAKKTIFVSHNIRQVREFCTKIAWIEGGLKEFGELEEVLPDYEAFKTFKKKSKAE
QKEFRNKLDSESRFVVK*

25 Sequence 631

Contig_0499_pos_6176_6922,

putative peptide of unknown function

atgattttattacaaaattttattactacaactatccaattaaatatctatcttatttta
gttattggactgctttacgtaatcatccactattatagaaataaagggtgtaaacgctttc
30 ttagatattttattaaattatataccgggtacttacacacgaatttggccacgtcttattt
aacaactcgtcgttggaaggccaaagatcttgcattgtgacaagccctagagaaaga
aaagtcacttcacaacaaggctatgcgattacacaatctaaaggatacttaggtcagttt
attacaactataggtgggtatcttatgccaccattgatgttttaactggattggtatct
attcactatcaatatccaagtataatttattactatataatttatttttatatattat
35 ttctttattacttcccgtaaactatcacctttgattgtcattatactcatctcaagttaa
ctctatttagtatttaaaacaagaccatcaatggttcatttacgacattgtcacattaagt
taccattttatttttaggcgtacttttaggtgaaatattacaatcctcatggacgattttt
cgtcttaccttcaacgacctaaaccttcttgggatggcagtgctttaaagaaagtact
cgagtacccacctttatctttagtttagtggtgatatattcaatctctatactgtgtat
40 ttattaatcaaatacacataactataa

Sequence 632

MIYLNQNFITTTIQLNIYLILVIGLLYVIIHYRNKGVNAFLDIYLNYPVLTHEFGHVLF
NKLGGKAKDLVIVTSPRERKVTSQQGYAITQSKGYLGQFITTTIGGYLMPPLMFLTGLVS
45 IHYQYPSIFITIYLFIFIIYFFITSRKLSPLIVIIILSSLLYLVEFKQDHWFIYDIVTSL
YHFILGVLLGEILQSSWTIFRLTFQRPKPSWDGSALTQVTRVPTFIFSLVWILFNLYTVY
LLIKYTIL*

Sequence 633

50 Contig_0499_pos_12156_11095,

putative peptide of unknown function

gtgataatcgacagattgcagagttatgttaccttatttggagagagtgccattccaaaaa
atgattttctaaaaatgatgaagataaagttactgagagtaaacctaagcgtagcttatat
gcactaatcatgactctatgtggtgtacatggaaccatttcactcgcaatcgcccttaacc
55 ttgccatattttattagcaaatcatgaaacatttgcttatcgaaatgatttattatttatt
gcttccggaatggtaatattaagtttaattattgcacaagtcattctgccttttagtaacg
cctgatagccctgaagtgaagataggtaatatgtcatttaagaggcgagaatctacatt
ttagaacatgttatcgattacctaataatcaaaaatcgacggttgaaacgagttaccgttat
ggaaacgtcattaagattaccacgataaaactcacatttttaaaaacggttgaaaaggaa

gatgaaaactccaaagaactagaacgacttcaaaagattgcatttaacgtagaaacaaaa
 acgctagagaaattggttgatgatggcgaaattactgagagtgttcttgaaaactatatg
 cgttacgctgaacgaacagaagtgatataaacaagcttcggttattaagacgaattattggt
 5 ggtttaagaggaatgctattgaaacgctcgtgtaaaaacaaaaattaattcggcatcatct
 cttagtgttactgataaacttattagaattgggtaaaatcaataagcttggttcattataac
 gtcgtaagtcggttagccaaagaagctactactgataataaactagaagtaggtatgatt
 tgcgatggatctctgatgagaatagataacttaacaccaaacaatttctttaattccaga
 catgaagatacacttaccaaaattaaattaaatgctttaagagaaacgcgtattcta
 10 agagaactaattgaaaatgatgagataacagaaggtactgcattaaaattaagagaatcc
 attaattatgatgaaatggttaattgtagatagtatgacataa

Sequence 634

VIIDRLQSYVTLFGESPFQKMISKNDKVTESKPKRSLYALIMTLGCVHGTISLAIALT
 LPYLLANHETFAYRNDLLFIASGMVILSLIIAQVILPLVTPDSPEVKIGNMSFKEARIYI
 15 LEHVIDYLNQKSTFETSRYGNVIKDYHDKLTLKTVKEKEDENSKELERLQKIAFNVETK
 TLEKLVDDGEITESVLENYMRYAERTEVYKQASLLRRIIVGLRGMMLKRRVTKINSASS
 LSVTDNLELGKINKLVHYNVVSRLAKEATTDNKLEVGMICDGYLMRIDNLTNNFFNSR
 HEDTLTKIKLNALREQRRILRELIENDEITEGTALKLRESINYDEMIVIDSMT*

20 Sequence 635

Contig_0499_pos_5935_5243,
 is similar to (with p-value 2.0e-31)
 >sp:sp|P27620|TAGA_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN
 A. >pir:pir|B49757|B49757 techoic acid synthesis protein tag
 25 A - Bacillus subtilis (strain 168) >gp:gp|M57497|BACTAGABCD
 2 B. subtilis tagA, tagB, tagC and tagD genes, complete cds.
 NID: gl43722. >gp:gp|Z99122|BSUB0019_72 Bacillus subtilis co
 mplete genome (section 19 of 21): from 3597091 to 3809700. N
 ID: g2636029.
 30 atgctggaaatggtagaaaatattaaacaattcatatctagcaatacagatgataattta
 ttttagtgactgctaactcctgaaatcgtggattatgcaactgaacatgagctatataga
 aatttaattaatcaagctgattatgtagttccagatggtacaggaatagtaaaagcttca
 aagcgattaaaacagcccttaaaacggcgtgtgccaggaatagaacttcttgaagaatgt
 ctgaaaatagcacatgtcagccatcagcgcgtatatctgcttgatctaaaaatgaaatt
 35 gttgagtcagcagagaaaaaacttcaatctcaataccctaataatccactttgcacatcat
 catggctatattcatctagaagatgaaacagtcataaaacgtataacaagttttaatccc
 gattacattttttaggaatgggatttccaaagcaagaacaatggattcaaaagcataag
 gacaagttaagcacactgtgatgatggcgtaggtgggtcgtttgaagtattcagtggc
 tcaaagaaaagcagcacctcaaatatttagaaagttaaatttgagtggttatatcgtgtg
 40 cttattgattggaaacgcattgggagaatgataagttattcctaaatttatgttaaaggta
 gcaatacaaaaatataaaatgaaatcaaaataa

Sequence 636

MLEMVENIKQFISNTDDNLFIVTANPEIVDYATEHELYRNLINQADYVVPDGTGIVKAS
 45 KRLKQPLKRRVPGIELLEELKIAHVSHQRVYLLGSKNEIVESAEKKLQSQYPNIHFAHH
 HGYIHLEDETIVIKRITSFNPDYIFVGMGFPKQEQWIKHKDKFKHTVMMGVGGSFEVFSG
 SKKRAPQIFRKLNIWVYRVLIDWKIRGMISIPKFMLKVAIQYKMKSK*

Sequence 637

50 Contig_0499_pos_3802_3227,
 is similar to (with p-value 3.0e-29)
 >sp:sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEAS
 E PROTEIN TAGG. >gp:gp|U13832|BSU13832_1 Bacillus subtilis 1
 68 highly hydrophobic integral membrane protein (tagG) gene
 55 and ATP-binding protein (tagH) gene, complete cds. NID: g755
 151. >gp:gp|Z99122|BSUB0019_68 Bacillus subtilis complete ge
 nome (section 19 of 21): from 3597091 to 3809700. NID: g2636
 029.
 atgtggttctttattaatcaaggtgtcctagaaggaactaaatcaatctcacagaatttc

aatcaagtggcaaagatgaatttcccactctcaatcattcctacttatattgtaacaagt
 aggttctatggtcatttaggattattagcaattattataatagcttgatgttcaatgga
 attatcccttcaattcacattgtacaattacttatatatgtaccttttgcatatttgcta
 acatcgtcggtggcacttttaacatccactttggggattttaattagagatacgagatg
 5 attatgcaagcattaatgagaatattgttttatatgtctccaattttatgggtgccaaaa
 aatcacggcgtaagtgggttgattcatcaaattatgttatttaatccagtatattttatc
 gcagaatcataccgagcagcgatattgttccatcaatggattttcatagatcattggaag
 ttaatgctatataacggttattatcattctcttattctttatagtaggttctattttacat
 agacgctatagagatcactttgcggaacttcttgtaa

10

Sequence 638

MWFFINQGVLEGTKSISQKFNQVAKMNFPLSIPTIYIVTSRFGHLGLLAIIIIACMFNG
 IIPSIHIVQLLIYVFPAYLLTSSVALLTSTLGILIRDTQMIMQALMRILFYMSPIWVVK
 NHGVSLIHQIMLFNPVYFIAESYRAAILFHQWYFIDHWKMLYNVIIILLFFIVGSILH
 15 RRYRDHFADFL*

Sequence 639

Contig_0499_pos_2865_1891,

is similar to (with p-value 4.0e-36)

20 >sp:sp|P27621|TAGB_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN
 B PRECURSOR. >pir:pir|C49757|C49757 techoic acid synthesis p
 rotein tagB - Bacillus subtilis (strain 168) >gp:gp|M57497|B
 ACTAGABCD_3 B.subtilis tagA, tagB, tagC and tagD genes, comp
 lete cds. NID: g143722. >gp:gp|Z99122|BSUB0019_73 Bacillus s
 25 ubtilis complete genome (section 19 of 21): from 3597091 to
 3809700. NID: g2636029.

gtgcttttaaatatgggttttaaacggtttaataataaattcgaagcacattgtgataatg
 atgacctttaagcaagatatactgcctattatagaggccttatgtagtgaaggataccat
 gtgacggttataggtaaaaaaatatatcagaaagatattaataacattaatcatgcatat
 30 tttataacctgccggaaataaatatattatgagacatatgaaagtattaagtaaagcaag
 gttattattttagatacgtattatttaaatgatgggtggctatcagaagaaaaaagggcaa
 actgtttattcaaacatggcatgctgctggcgcttaaaaattttggcttaactgatcat
 caagttgatttaaaaaataaggctatggtaagacaatacaaaaaagtttatgatgctacc
 gattattatttggtaggtggggagaaaatggctcaatgtttataacaatcgtttgatgca
 35 tctccatcgcaaatgttaaagtttgacttccaagactgacccaatactttagaagcaat
 ctttaagttagaacaacaacgattaaaaaagaaatatcatattacaataaaactcgcaqta
 tatgttccgacttatagagaaggtcaagtagcacacgtactattgataaagaaaacttt
 gaacggcacttgccgaattatacgttatttgagtcatttgcaccttcgactgttgattgt
 caaacttctcattcaatcgatgttacttcattgttaattatggcgatattattataagt
 40 gattatagctcattacattgaagcaagcgacttaataaaccgacactattttataat
 tatgatgaacgcaatatgaaaaagtaagaggattgaatgaattttattatgctattcca
 gaacgatacaaaatgagtaatgaagagtcattatacaagcgatacaggataacgatgag
 caatttcaatcttag

45 Sequence 640

VLLNMVFKPFNINSKHIVIMMTFKQDILPIIEALCSEGYHVTVIGKKIYQKDINNINHAY
 FIPAGNKYIMRHMVLSKAKVILDTYYLMMGGYQKKKGQTVIQTWHAAGALKNFGLTDH
 QVDLKNKAMVRQYKKVYDATDYLVGGEKMAQCFIQSFDASPSQMLKFGLPRLTQYFRSN
 LKLEQQRLKKKYHITNKLAVYVPTYREGQVAQRTIDKENFERHLPNYTLLSHLHPSTVDC
 50 QTSHSIDVTSLLIMADIIISDYSSLPIEASALNKPTLIYNYDEQQYEKVRGLNEFYAIP
 ERYKMSNEESIIQAIQDNDEQFQS*

Sequence 641

Contig_0499_pos_1878_1240,

55 is similar to (with p-value 4.0e-66)

>gp:gp|AF008219|AF008219_3 Borrelia afzelii R-IP3 chromosome
 right end, arcA and arcB genes, complete cds. NID: g2697111

atgaaaaatttacgtaacagaagcttttttaactttatttagacttttcacgacaagaggta

gaattttttattaacactctccgaagatttgaagcgtgccaaatatatcggcactgaaaag
 cctatgctaaaaataaaaaatatacgcgcttctttttgaaaaagattccactagaacacgt
 tgcgcattcgaagttgccgcacatgatcaaggtgcacacgtcacttatcttggacctaca
 5 ggttctcaaatgggtaaaaaaagaaactgctaaagatacagcacgtgtacttgggtggtatg
 tatgatggattgagtagcagggtttctctcaacgtactgtagaacattagcgcaatat
 tcaggtgttccgggtatggaatggattaaccgatgaagatcacctacacaagtgttgc
 gactttttaactgctaaagaagtattgaaaaaagagtatgctgatatcaactttacttat
 gttggcgtggacgtaacaatgttgctaacgcattaatgcaaggtgctgccattatgggt
 10 atgaatttccatcttgtttgtcctaaagaactcaatccgacagaagaattattaaatcgt
 tgcgcagctattgcgcaggaataatggcggtaacattttaa

Sequence 642

MKNLRNRSFLTLDFSRQVEVEFLTLSEDLKRAKYIGTEKPKMLKNKNIALLFKEDSTRTR
 CAFEVAHHDQGAHVITYLGPTGSQMGKETAKDTARVLGGMYDGEYRGFSQRTVETLAQY
 15 SGVPVWNGLTDEDHPTQVLADFLTAKEVLKKEYADINFTYVGDGRNNVANALMQGAIMG
 MNFHLVCPKELNPTEELLNRCVLRKMAVTF*

Sequence 643

Contig_0500_pos_5053_3860,
 20 is similar to (with p-value 0.0e+00)
 >sp:sp|Q07908|ARGJ BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2
 .3.1.35) (ORNITHINE ACETYLTRANSFERASE) (ORNITHINE TRANSACETY
 LASE) (OATASE) / AMINO-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-
 ACETYLGUTAMATE SYNTHASE) (AGS). >gp:gp|L06036|BACACETYL 2
 25 Bacillus stearothermophilus ornithine acetyltransferase (ar
 gJ) and acetylglutamate kinase (argB) genes, complete cds's,
 argC gene, 3' end, and argD gene, 5' end. NID: g304133.
 atgaatataattaagggaaatattgcaagtcctcttggattttcagctgatggctgcac
 gctggcttttaaaaagaaaaaattagacttttgggtggattgtttcagaagtacctgcaaat
 30 gtagctggtgtattttacaactaataagggtcattgctgcaccattaaaattaacaaaaaac
 agcatcgaaaaaagtggttaaaatgcaagctattgttgtaattcaggtattgctaattct
 tgtactggtaaaacaaggagaaaaagatgcttttaaaatgcaacaactggccgcaataaa
 ttacaaattcaaccagaatatgttggtgctgcacatctactgggtgtatttgaaagggtgatg
 ccaatgtctatttctaagaatggcttttccaaactagttaaaaaacggtaattgctgatgac
 35 tttgcaaaagcgatattaacaacgggatactcatacaaaaacatgcgttgtaaacgaagaa
 tttggtagcgatacagtaacgatggcaggtgtagcaaaagggcaggaatgatacatcct
 aatttggtacaaatgctagcatttataacctgtgacgctaacatctcatcaaaacatta
 caacaggctttaaaagatgtggttgaaagttacattcaatcaaatcactgtagatgggtgac
 acttcaacaaatgatattggtgcttgtgatgtcaaatggatgtacaaataataacgaaatt
 40 aaaaaagacagcgaagactactataaatttaagcagatgcttctatatattatgaccgat
 ttagcaaaaagtattgcaagggatggcgaagggtgcttctaaattaatagaagtcacgggt
 aaagggtgcaaaagaatctagtgtgcaagaatgattgctaaaagtgtggtgggttcaagt
 ttagtaaaaaccgcaatttttggcgaagatcctaattggggtagaattattgctgctgca
 45 ggttatgctaaaacataattttgatattaatcaggtagacatttttataggtaggatacct
 gtattaataagatcctcaccagtaaaagtacgataaagaagaattcaagaaataatgagt
 gctgaagaaatatcaattcagcttgaccttcatcaagggaattgtgaaggtcaagcatgg
 ggtgtgatttatcgtagctacgttaaaatcaacgcactatacaccacttag

Sequence 644

MNIIKGNIASPLGFSADGLHAGFKKKKLDGFWIVSEVPANVAGVFTTNKVIAAPLKLTKN
 50 SIEKSGKMQAIVVNSGIANSTGKQGEKDAFKMQQLAANKLQIQPEYVGVASTGVIGKVM
 PMSILKNGFSKLVKNGNADDFAKAILTDTHTKTCVVNEEFGSDTVTMAGVAKGSGMIHP
 NLATMLAFITCDANISSQTLQALKDVEVTFNQITVDGDTSTNDMLVMSNGCTNNNEI
 KKDSEDYKFKQMLLYIMTDLAKSIARDGEGASKLIEVTVKGAKESSAARMIAKSVVGS
 55 LVKTAIFGEDPNWGRIIAAAGYAKTYFDINQVDIFIGRIPVLIRSSPVKYDKKEIQEIMS
 AEEISIQLDLHQGNCEGQAWGCDLSYDYVKINALYTT*

Sequence 645

Contig_0500_pos_1725_1258,

is similar to (with p-value 5.0e-27)

>sp:sp|P49786|BCCP_BACSU BIOTIN CARBOXYL CARRIER PROTEIN OF ACETYL-COA CARBOXYLASE (EC 6.4.1.2) (BCCP).

5 atgaactttaaagaaataaaagaattaatcgaaattcttgatcaatctagtttaactgaa
ataaatattgaagataataaaggtagcgtagttaatttaaaaaaagaaaaagagactgaa
atagttacaccgcaagttactcaacaaccaactcaaccgataaatcatacgcataatgaa
acacaacaaaagccatcacatagctctaaagatgaacaaagtagtgataatgaatacaat
accattaatgcaccaatggttggtacattttataaatcaccttcaccagatgaagaagca
tacgttcaagttggagataaaagttacgaatgaaagtagtgtttgtatattagaagctatg
10 aaattatttaatgagattcaagccgaaacaacaggtgaaatcatagaaatttttagtagaa
gacggacaaatggtagagtatggccagccgttatttaaggtgaaataa

Sequence 646

15 MNFKEIKELIEILDQSSLTEINIEDNKGSVVNLKKEKETIIVTPQVTQOPTQPINHTHNE
TQQKPSHSSKDEQSSDNEYNTINAPMVGTFYKSPSPDEEAYVQVGDKVTNESTVCILEAM
KLFNEIQAETTGEIIEILVEDGQMVEYQGQLFKVK*

Sequence 647

Contig_0500_pos_0_925,

20 is similar to (with p-value 6.0e-95)

>sp:sp|P49787|ACCC_BACSU BIOTIN CARBOXYLASE (EC 6.3.4.14) (A
SUBUNIT OF ACETYL-COA CARBOXYLASE (EC 6.4.1.2)) (ACC). >gp:
gp|U36245|BSU36245_2 Bacillus subtilis biotin carboxyl carri
er protein (accB) and biotin carboxylase (accC) genes, compl
25 ete cds. NID: g1055244.

atgggaataaaagatattgctaaagactgaaatgattaaagccaatgtacctgtagtacca
ggaagtgaaggacttattcaaagtatagatgacgctaaaaaaatagctaaaaaaatcggc
tatccagttatcatcaaagccacagcaggtggtggtggaaggtattcggttgctcgt
gatgagaagaacttgaaactggttaccgtatgacacaacaagaagctgaaaccgcgttc
30 ggaaatggtggtttatacttagaaaaatttatagaaaaacttttagacatatagagattcaa
attattggcgatacttatggaaacgttatatacttttaggtgaacgtgattgtacaattcaa
agaagaatgcaaaagctcgttgaagaagcaccctcaccagttttaagtgaagataaacgc
caagaaatgggtaattgctgcaatttagagccgcaaaagctgtaattatgaaaacgcaggt
acaattgaatttatatatgatttagatgataaccaattttatttcatggaaatgaataca
35 cgtattcaagttgaacacccagtaactgaaatggtaacaggagtagatttagtaaaatta
caactcaaagttgctatgggtgaggcggttaccttttaacaagaagatatttccattaac
ggtcacgctattgaatttcgaatcaatgctgaaaaatccttacaaaaactttatgccatca
ccaggcaagattacccaatatcttgctccaggcggttttgagtgagaattgaatcagca
tggtatactaattatagataccaccttactatgactccatggtggcaaaacttatagtt
40 cagcaacctacacgtgaagaatcaattatgacaggcattcgtgctttaagtgaatatctt
gttttaggtatcgacactaTGATTT

Sequence 648

45 MGIKDIAKAEMIKANVPVPGSEGLIQSIDDAKKIAKKIGYPVVIKATAGGGGKGIRVAR
DEKELETGYRMTQQEAETAFGNGGLYLEKFIEFRHIEIQIIGDTYGNVIHLGERDCTIQ
RRMQKLVEEAPSPVLSEDKRQEMGNAAIRAAKAVNYENAGTIEFIYDLDDNQFYFEMMNT
RIQVEHPVTEMVTGVDLVKLQKLVAMGEALPFKQEDISINGHAIEFRINAENPYKNFMPS
PGKITQYLAPGGFGVRIESACYTNYTIPPYDSMVAKLIVHEPTREESIMTGIRALSEYL
VLGIDTMIX

50

Sequence 649

Contig_0501_pos_9189_8275,

is similar to (with p-value 0.0e+00)

55 >gp:gp|U94706|SAU94706_2 Staphylococcus aureus strain ATCC 8
325-4 cell wall/cell division gene cluster, yllB, yllC, yllD
, pbpA, mraY, murD, divlB, ftsA and ftsZ genes, complete cds
. NID: g2149889.

atgttaaaccgaaaccattgattattttaaatattaaagaagatggtgtgtatgttgactgt
acgttgggtggagcaggacatgccctctatttacttaataattaaatgataaaggtaga

cttattgcgattgatcaagatttaacagccatagaaaatgcgaaagaagttttaaaaagaa
 catttgcacaaagtcacttttgttcataacaactttcgagaattaacaaatattttaaat
 gaattagaaattgaaaaagtagatggattttattatgacttaggtgtttcaagccgcaa
 ttggatgtgcctgaaagaggcttttagttatcacaaatgatgcgaaactagatatgcgaatg
 5 gatcaaacacaatcacttttctgctatgaagtagttaatcaatggctttatgaagcatta
 gttaggattttctttcgttacggtgaagagaaaattttctaaacaaattgcacgcagaatt
 gaagcccatcgagaacaacaacctatagaaacaactttagaactagttgatgtcattaaa
 gaaggcataccagcgaaagcaagacgaaaaggggacatcctgcgaaacgcgtgttccaa
 gctattcgaattgctgtgaatgatgagttatcagcttttgaagattcagttgagcaagcc
 10 attgaatgtgtgaaggtcggaggtagaatttcagttattactttccactctttggaagat
 cgtttgtgtaaaacaaattttccaagagtttgagaaaggtccagacgtaccaagaggtctc
 ccggtattcctgaagcatatacacctaagttaaaacgagtaaatcgtaaacgattacc
 gctactgatgacgatttaaacgaaaacaatcgagcacgtagcgccaagttacgcgtagca
 gaaatattaaaataa.

15

Sequence 650

MLNETIDYLNKEDGVYVDCTLGGAGHALYLLNQLNDKGRLIAIDQDLTAIENAKEVLKE
 HLHKVTFVHNNFRELTNINLELEIEKVDGIYYDLGVSSPQLDVPERGFSYHNDAKLDMRM
 DQTQSL SAYEVVNQWSYEALVRIFFRYGEEKFSKQIARRIEAHREQQPIETTLELVDVIK
 20 EGIPAKARRKGGHPAKRVFQAIRIAVNDELSAFEDSVEQAIECVKVGGRISVITEHSLED
 RLCKQIFQEFEGKPDVPRGLPIPEAYTPKLKRVNRKPITATDDDLNENNRARS AKLRVA
 EILK*

Sequence 651

25 Contig_0501_pos_7793_5553,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AB007500|AB007500_3 Staphylococcus aureus genes for p
 enicillin-binding protein 1, MraY, MurD, partial and complet
 e cds. NID: g2463558. >gp:gp|U94706|SAU94706_4 Staphylococcu
 30 s aureus strain ATCC 8325-4 cell wall/cell division gene clu
 ster, yllB, yllC, yllD, pbpA, mraY, murD, divlB, ftsA and ft
 sZ genes, complete cds. NID: g2149889.
 gtgctaaggttttcttatgtaatgattactggccactctaattggtaagatttaattatg
 aaagccaatgagaaatacttagtcaaaaaattctcaacaaccagaacgaggttaagatttac
 35 gatcgtaatggtaaggttttagcagaagatgtagaagatataaacttggtgcagttgtg
 gataaaaaagcaagtaaaagaaagtaaaaaagccgcgacacgtggttgataaaaaaaagaca
 gcaaaaaaatttagctgaaatcatagatatggacgctgacgaaatagaaaaacgacttaat
 aataagaaaagcctttcaaatcgaatttggtcagaaaggtaactaatttaacttatcaagaa
 aaagaaaaaatagagaaaaatgaaattacctggtatagcactttaccagaaactgagcga
 40 ttttatcctaattggtaattttgcctcccatttaatagggatggcacagaaagatcctgat
 actggtgaattaaatgggtgcattaggtgttgaaaaaatatttaatatgttatttaaatgga
 tcaagaggtgcacttaaatatatacatgatatatggggctacatcgacctaatacgaag
 aaagagcagcaacctaaacgtggagatgatgtacacttaacaattgattctaataatacaa
 gtctttgtggaagaagctcttgatgacatggtgaacggtatgctccaaaagatttattt
 45 gcagtagtaatggacgcaaaaactgggtgaaatacttgcatatagccaacgtccaactttt
 aatcctgaaaacaggtaaagattttggcaaaaagtgggcgaacgatttatatcaaaaataca
 tatgaaccgggctctacttttaaaacatacggcttagctgcagcaattcaagaaggtaaa
 ttcaaaccggatgaaaagtataaatcaggtcatagaaatattatgggctctgaaatttcc
 gattggaataaaaactggttggggacgtatacctatgtcggttaggttttacttattcatca
 50 aatacgttgatgatgcacttaagaatttgggttggtgcccataaaatgaaatcttggtat
 gaacgctttggatttggcaaaaaacgggtggtatggttgatggagaagctgcaggtaat
 attggttgggcaaatgaattacaacaaaaacgtcagcatttgggtcaatccacaactggt
 accctgtcacaatgattcaagcacaatcggtcttctttaataaaggaaatagcttaaa
 ccatggtttgtaagtagtattgataatccaataactaaaaagaattattactctggtaaa
 55 aaagagtttgtcggtaaaccagtaacggaagaacagccaataaagttgaagaagaactt
 gataaagtagtaaatagtaagaagagtcagctatgaattatcgcgtaaaaggttatgat
 attgaaggttaagacaggaacagcacaagtagctgattcaaatggaggcggttatgttaaa
 ggtgaaaatccttactttgtaagcttcatgggggatgcacctaagaaaaatcctaaagtc
 attgtctatgcaggtatgagcttctgctcaaaaaaatgatcaagaagcatatgaaatgggt

gtgagcaaagcattttaaccaattatggagaatacgcctgaaatattttaaatgttgaaaa
 tctagtgtacttcatcaaaaactgactatagtaaagtcctaactgcaaggagatgaa
 gttcaaaaagcagaggatagcgtcaatgctcaatctcttaaacctattacgattggtaat
 ggcaaacagattaaacaacaatcagtttaagtcaggtaccaaaagtcctaccacacagtaaa
 5 gtaatgttaatgacagacggggaattaacaatgccggatgaccggatggacaaaggaa
 gatgtacttgcttttgaagatttaacgaaaacttaaagtttctactaaaggtaatggattt
 gtcacgaatcaaagtatctcaaaaggtaaatcattaaaaataaagataagatagaagt
 tcattatctgtcgaagatacggatgatgaccaagagaaaactgatgaggactcttcggat
 acaaatcaaagaagataaagctgatgaggatcattcaatacatcttcgtcaactaag
 10 aatgataagtcaaacgccgactcgaaaaatgattctgatgacagcacaatgaaacatca
 ggttctgagagaaataattaa

Sequence 652

VLRFSYVMITGHSNGQDLIMKANEKYLKNSQQPERGKIYDRNGKVLAEVERYKLVAVV
 15 DKKASKESKKPRHVVDKKKTAKKLAEIIDMADEIEKRLNNKKAQIEFGQKGTNLTYQE
 KEKIEKMKLPGIALYPETERFYPNGNFASHLIGMAQKDPDTGELNGALGVEKIFNSYLN
 SRGALKYIHDWGYIAPNTKKEQQPKRGDDVHLTIDSNIQVFVEEALDDMVERYAPKDLF
 AVVMDAKTGEILAYSQRPTFNPETGKDFGKKWANDLYQNTYEPGSTFKTYGLAAAIQEGK
 FKPDEKYKSGHRNIMGSEISDWNKTGWGRIPMSLGFTYSSNTLMMHLQDLVGADKMKSWY
 20 ERFEGKKTGGMFDGEAAGNIGWANELQOKTSAFGQSTTVTPAQMIQAQSAFFNKGNNMLK
 PWFVSSIDNPITKKNYSGKKEFVGKPVTEETANKVEEELDKVVNSKKSHAMNYRVKGYD
 IEGKTGTAQVADSNNGGYVKGENPYFVSFMGDAPKKNPKVIVYAGMSLAQKNDQEAYEMG
 VSKAFKPIMENTLKYLNVGKSSDTSKTDYSKVPNVQDEVQKAEDSVNAQSLKPITIGN
 GKQIKQQSVKSGTKVLPKSKVMLMTDGLTMPDMTGWTKEVDLAFEDLTKLVSTKGNF
 25 VTNQSISKQIIEKNKDKIEVSLSAEDTDDQEKTDDESSDNKSKDKADEHDHNTSSSTK
 NDKSNADSKNDSDDSTNETSGSERNN*

Sequence 653

Contig_0501_pos_5286_4399,
 30 is similar to (with p-value 0.0e+00)
 >gp:gp|AB007500|AB007500_4 Staphylococcus aureus genes for p
 enicillin-binding protein 1, MraY, MurD, partial and complet
 e cds. NID: g2463558.
 atgaagtttgacaaagtatccgtgaggaaggcctcaaagccatatgaaaaaacaggt
 35 actcctactatgggtgggcttacatttttaattagtattataattacctctatcattgca
 attatctttgtagaccattcaaatccaattattttgttactatttgaacaatcggtttt
 ggtcttatttgatttattgatgactatattattgtagttaaaaagaataaccaaggatta
 actagtaaaacaaaagtttctagcacaataattattgcagttatattctttgtgctaagt
 gatgtatttcaccttgtgcattttacgacagatttgcataattccatttgtgaattttgat
 40 attccggttgctatttgcattatgtgatatttatcgctcttttgcaagttggtttctcaa
 gctgtaaacttaactgatggttttagatggattggcaactggtttgtcaataataggtttt
 gcaatgtatgctgtaatgagttacatggttagattcaccggctattggcatattttgtatt
 ataatgattttcgctttactaggtttcttaccttacaatttaaatccagcgaaagttttc
 atgggagacacaggaagtcttgccttaggtggtatttttgcaacgatttcaatcatgttg
 45 aatcaagaattatcattaatattaattgggtttgtgtttgtagttgagacattatctgta
 atgttacaagtagcctcatataaattaacgaagaaacgtattttcaagatgagtcctata
 catcaccacttcgaattaaagtgggtgggtgaatggaaagtagtaacagatttttggacg
 gtaggtttaattacgggattaatagggtttatggattggagtgcattaa

Sequence 654

MKFGQSIREEGPQSHMKKTGTPTMGGLTFLISIIITSIIAIIIFVDHSNPIILLFVTIGF
 GLIGFIDDIYIVVKNNQGLTSKQKFLAQIIIAVIFVFLSDVFLVHFTTDLHIPFVNFD
 IPLSFAYVIFVFWQVGFSAVNLTDLGLATGLSIIIGFAMYAVMSYMLDSPAIGIFCI
 IMIFALLGFLPYNLNPAKVFMDGTGSLALGGIFATISIMLNQELSLILIGFVVFVETLSV
 55 MLQVASYKLTKKRIFKMSPIHHHFELSGWGEWKVTVFWTVGLITGLIGLWIGVH*

Sequence 655

Contig_0501_pos_4235_3048,
 is similar to (with p-value 0.0e+00)

>gp:gp|AF009671|AF009671_1 Staphylococcus aureus UDP-N-acetylmuramoyl-L-alanine : D-glutamate ligase (murD) gene, complete cds. NID: g2305091.

5 atgggcattgaggttaattagcggtagtcatcctttttctttatttagatgatgatcctatc
attgtgaaaaaccagggtattccatatactgtatcaattattaagaagcagcaaataga
gggcttaaaatcttaacagaggttgaacttagctatttaatttctgaggcaccaatcata
gcagttactggaactaacggtaaaactactgtcacttcactaatcgggtgatattttccaa
aaaagcgtgttgactggacgactttctgggaatattggttatgtagcctcaaaagttgca
caagaagttaaatcagatgagtatttaataacagaattatcatcttttcaattattaggc
10 attgaagaatataaaccacatatcgctatcattactaatattttattctgcacatttgat
taccatgaaacgtagagaactatcaaaatgctaaaaagcaaatatataaaaatcaaact
aaagatgattatctcatttgaattatcatcaagacacctaattgaatcagaaaatcta
gaagcgaacacataaatttttttcaacacagcaagaagttgatgggatatacattaaagat
ggtttcattgtttttaacggcattcgcatattaacactaaagacttagtgctaccagga
15 gaacataacctggaaaatatttttagcagctgttctagcatcaatcattgctggagtcca
gtcaaagctattgtagatagctgttacttttccgggtattgatcatagacttcagtat
attggtacaaatcgcacaaataaattataatgattcaaaagcaactaatcttttagct
actcaatttgcgcttaactcttttgatcaaccaattatttggttggtggtggttagat
cgtggtaatgaattcgatgaacttattccttatatggaaaatgtacgtgtgatggttgtt
20 tttggagaaacacagaataaatttgctaaattgggaaatagtcaaggtaaagtatgtgatt
aaagcaacagatgtagaggatgctgttgataaaaattcaagatatagtcgagccaaatgat
gttgttctattatcaccagcttggtgcaagttgggatcagtatcatacatttgaagaacgt
ggtgagaagtttatcgatagattccgagcgacttgccatcatactaa

25 Sequence 656
MGIEVISGSHPSLLDDPIIVKNPGIPYTVSIIKEANRGLKILTEVELSYLISEAPII
AVTGTNGKTTVTSLIGDIFQKSVLTGRLSGNIGYVASKVAQEVKSDEYLITELSSFQLLG
IEEYKPHIAITNIYSAHLDYHETLENYQNKKQIYKNQTKDDYLICNYHQRHLIESEN
EAKTFYFSTQQEVDGIYIKDGFIVFNGIRIINTKDLVLPGEHNLENILA AVLASIAGVP
30 VKAIVDSLVTFSGIDHRLQYIGTNRTNKYYNDSKATNTLATQFALNSFDQPIIWLCGGLD
RGNEFDELIPYMNVRVMVVFGETQDKFAKLNSQGGKYVIKATDVEDAVDKIQDIVEPND
VLLSPACASWDQYHTFEERGEKFIDRFRAHLPSY*

Sequence 657

35 Contig_0501_pos_3038_1620,
is similar to (with p-value 3.0e-89)
>gp:gp|U94706|SAU94706_7 Staphylococcus aureus strain ATCC 8
325-4 cell wall/cell division gene cluster, yllB, yllC, yllD
, pbpA, mraY, murD, divlB, ftsA and ftsZ genes, complete cds
40 . NID: g2149889.
atgctcatggaagagataaaaaatcaacctaataaggagaatatgtcgaataaagacgat
aatgcaactcatttgaatgacagtcacagaaatgaagatttagagcttttttagacggaat
aaaaacgctcgccaacgcagaagacgctcgtatagataaccaaagtaagaaaaggatgct
acgtctacacaatcacagttagaaactaaaccaatggataaatttttagataatcacaag
45 tcgcataatcaaaacaaagaaataaaaagtacttaattgaagaaaatgttaatgatgaa
aacgacaatcaaaaaatattaatgataaattaaatgaccgtattgtccaagaacaaat
gaaagtcgtcaaaagtactgaagacgatgaggaaatttctacttgatcatcgagtgaaaca
caacctaaagcctctcgtcattctaaaaagcataaattactaagtaaatcttctctaaa
aaagaaaaggaaacatctacatcggttcaatagtaatgagaaggtaactcaaattaaaccg
50 cttagtttagaagaaaaagagccataagacgtaaaaagcaaaaagaatccaatatacc
attatcacactactcattcttatcattgttctcattttactctatatgtttacaccactg
agtaaaatatcaaatgtaaatgttaaaggtaataacaacgtaagtacgagtaaaataaag
aaagaacttaacgttacttcgcatcacgaatgtatacttttagtaaaaataaagcgatt
aggaacttaaaacagaatccttttaatacaagaagttgatattcataaacaattaccaaac
55 acgttaactgtcaacgtgactgagtaaccaattgtcgggttagaaaaaaataaagataaa
tatgtgccaatatagaagatggttaaagaattaacagaatacaaaagatgaagtgtcacat
gatgggcctatcattgatggtttcaaaggagacaaaaaacacgaattataaaagcttta
tcagaatgtcacctaaagtgaagaaacttaattgcagaggtgagttacgcaccaactaaa
aataaaacaaagtcgcataaaaatcttcaccaagataaatatgcaagttattggtgacatt

acaacgattgcagacaaaatgcaatattatcctcaaatgtcacaatcattaagcagagat
 gactctggcgaaacttaagacaaaatggctatatattgatttatcggttgagcgctcatttatt
 ccttatcaagggtcatcaactgttcaatcggttacagaacaaaatgtaaccaagtcaaca
 caagaagaaaatgatgcaaaagaagaacttcaaaatgtgttgataaaaataaataaaca
 5 tctaccagtggcgaggcgactttctggtctgtaactga

Sequence 658

MLMEENKNQPNKENMSNKDDNATHLNDSHRNEDLELFRNKNARQRRRRRIDNQSKEKDA
 TSTQSQLETKPMDKFLDNHKSNNQKEIKSDLIEENVNDENDNQKNINDKLNDRIVQETN
 10 ESRQSTEDDEEFLDHRSEQQPKASRHSKHKLLSKFTSKKEKETSTSFNSNEKVTQIKP
 LSLEEKRAIRRKQKRIQYTIITLLILIIIVLILLYMFTPLSKISNVNVKGNVNVSTSKIK
 KELNVTSRSMYTFSKNKAIRNLQNPLIKEVDIHKQLPNTLTNVNTEYQIVGLEKNKDK
 YVPIIEDGKELTEYKDEVSHDGPIDGFKGDKKTRIIKALSEMSPKVRNLIAEVSYPATK
 NKQSRIKIFTKDNMQVIGDITTIADKMQYYPQMSQSLSRDDSGELKTNGYIDLSVGASFI
 15 PYQGSSTVQSGTEQNVTKSTQEENDAKEELQNVLNKINKQSTSGEGDFLCN*

Sequence 659

Contig_0501_pos_1207_863,
 putative peptide of unknown function
 20 gtgacaaacggaggaaggtggggatgacgtcaaatcatcatgcccttatgatttgggc
 tacacacgtgctacaatggacaatacaaaagggtacgaaaccgcgaggtcaagcaaatcc
 cataaagttgttctcagttcggttagtctgcaactcgactatatgaagctggaatcg
 ctagtaatcgtagatcagcatgctacggtgaatacgttcccggtcttgtagacacccgcc
 cgtcacaccacgagagtttgtaacaccggaagccggtggagtaaccatttgagctagcc
 25 gtcgaaggtgggacaaatgattgggggtgaagtcgtaacaaggtag

Sequence 660

VTNRRKVGMTSNHHPYDLGYTRATMDNTKGSETARSSKSHKVVLSDDCSLQLDYMKLES
 LVIVDQHATVNTFPGLVHTARHTTRVCNTRSRSWSNHLELAVEGGTNDWGEVVTR*

30

Sequence 661

Contig_0502_pos_1097_1513,
 putative peptide of unknown function
 atgcatgttgctaccaggcagcaaaaataacatatcaaaatcattcctataagatacac
 35 tttgataatggacatattgttcactctcactactgaaccgattattgcaactggatttgat
 gttaccacaaaaccctttaatagaacaactatttcaagtacgacaatcagaagttcaatta
 acagaattagacgaatctacaaagtttcctaatgtatttttaattggggcgactgtacgt
 catcaaaatgccattctttgttatatatataaaatcagagcacgttttgagtagtagca
 cgcatagtaagcctacgcgaaggttacctgaagatacatcattaattcagtcgtatcgt
 40 caaaaaaatatgtttctagacgattatagttgtgtgatgtgaattgcacatgttaa

Sequence 662

MHVGYQARKITYQNHSYKIHFNDGHIVHSHTPEIIATGFDVTQNPLIEQLFQVRQSEVQL
 TELDESTKFPNVFLIGATVRHQNAILCYIYKFRARFAVLARIVSLREGLPEDTSLIQSYR
 45 QKNMFLDDYSCCDVNCTC*

Sequence 663

Contig_0502_pos_1702_2889,
 putative peptide of unknown function
 50 atgattaaaaagttagagacttttgtaaggagagcaaaagcataccatcctcaatttcac
 aatattatttggatgggtgctttattgtacttatgtttgctttacccatctacttgtct
 tatcacgtttcagacggattacaatactacgtgtcacactggttaactaaactatcccaa
 atatcattatttcaagaaaataagctacaacataattttattcggtcattatggtgtcatt
 tcttttaggtacataattcattcgtctggcggttaccagttgttttcatgattagtttatct
 55 acagctcttatagatataactcatttaagcattatatcggtttggtctatcgaacctacg
 atgatgagacttggctcttcagatcagatatcatacccttgtagaaggatttggtatgt
 aatgctgctgctattactcaagctacacatcaatgtcatcggtgtacaaaagtacaatgt
 atgagcttggttaagtttcggaactgcatgtagttatcaaattggtgctacattatcgata
 ttcaacgcaagccaccgctcttggttggttttgcatacataggtatgggtttcttagga

ggaatcacataacaaactatggtatagtcataaacacctatgacaacacagtctgtt
 tttcaacgacaacctgtacgttgcccaaaccagctactcttaaaagcagcgtgaaa
 agtattcaaatgtttattgtacaagccttacctattttataggaattgccttattgta
 agtctattgtctcttacgtctattttgacttttatcaaatgcattcatacctttatta
 5 tggctactagatgtacctacacagcttgaccaggtattctgttttcaatgatacgtaa
 gatgggatgttggtggttaatatgaatggcggtactttaattcaaagactttccgcattc
 caattattgttgctagtcttttttagttcaacatttacagcatgttcagtaacaatgact
 atgctcatgcgtcgactcggttcaattctaggaattaaaatgataatgaacaaatggta
 10 tcgtccacaatttgcgtcacccatactagccatagcaatgttaagcataactaaaattca
 gacttaggagtgtgttatggaaatcattattatcggtgggttttttag

Sequence 664

MIKKLETfVKESKAYHPQFHNIICGWCFIVLMFALPIYLSYHVS DGLQYYVSHWLTkLSQ
 ISLfqENKlQHILFGHYGVISLgTYSFVWALPVVFMISLSTALIDITHLKHIVWSIEPT
 15 MMRLGLHESDIIPLLEGFGCNAAAITQATHQCHRCTKVQCMSLVSFGTACSYQIGATLSI
 FNASHRSWFLPLPYIGMVFLGGIIHNKLWYSHQTPMTQSVFQRQPVVRWPKPKLLKAAWK
 SIQMFIVQALPIFIGICLIVSLLSLTSILTFISNAFIPLWLLDVP TQLAPGILFSMIRK
 DGMLLFNMNGGTLIQRLSAFQLLLLVFFSSTFTACSVTMTMLMRR LGSILGIKMIMKQMV
 SSTICVTILAIAMLSITKISDLGVMLWKSLLSVVF*

Sequence 665

Contig_0502_pos_2964_3740,
 putative peptide of unknown function
 atgaatgaatttggaaaaagaagtgttgatggccaacttatagaacatcctgaagtacct
 25 atgagtgaatcactgaaggatgcatttgttgctgcgatgaaatcagacgtatcacacaa
 ctacatgaactatcttataaatatcaaccagatatcatctttattgaatgcagtgggtga
 gctgaaccactagctgtcgctcgatgcattcttcacacccgtacttgcaccttttatcact
 ttaaggagtatgggtgggaattattgatgcaagcatgtattcacgaattaaatcttatcca
 caagacattgcagctctattttatgaacaacttcgctattgttcgactttatttgtaaat
 30 aaaatagataagatagaggtggaagaaaccgcccgttgctacgtcaactcgagcgtctc
 aatagcgtatgccaatattcaagttgggtcaatttggagaattaaatttaaaatcactgcta
 gagccaacacatataaattcaaatgcattgtggcactttgcataagtaataataaatcatcaa
 ttcatcgaaaatcctaggctacaaacaaaagaagaatgattagtgcttagataacttg
 cctcaagatgtttaccgtgtcaaaagggtttgttcgtttttcagatcagcaacacgtttat
 35 ttagtacagtatgcacaaggaaatatagaattatctccattcaacttaaaaacgatgta
 ccattgtacctcattgttataggaaaacatttaaaacaaatataaatttgatttataa

Sequence 666

MNEFGKRSVDGLIEHPEVPMSEITEGCICCAMKSDVSQQLHELYLKYQPDII FIECSGV
 40 AEPLAVVDAFTFPVLAFFITLRSMVGIIDASMYSRIKSYPDIAALFYEQLRH CSTL FVN
 KIDKIEVEETARLLRQLERLNSDANIQVGQFGELNLKSLLEPT HINSNACGTLHSNINHQ
 FIENPRLQTK EEMISALDNLPQDVYRVKGFVRFS DQQHVYLVQYAQGNIELSPIQLKNDV
 PLYLIVIGKHLKQIQFDL*

Sequence 667

Contig_0502_pos_7308_8495,
 putative peptide of unknown function
 atgaagaaaaaattaagttatatgattaccattatgcttgccttttacgctaagtttagca
 cttggcctatttttcaatagtgtcacgccgactcgttaccacaaaagaatgggtgcaaac
 50 caaaaaacaactaaagtcactgtcagtaataaagacgttccagatgcagttcgcaaacctt
 gctgaagaaacaatatttatctcgtgtagctttattagataaaagcttccaaccacaaagca
 acatcgtatacacttgggtgaaccttttaaaattataaatttaataaggaaagcgacggc
 aattattattatccagtgtcataaaaaaggagatgtcgtttatgtagtaacaatttct
 cctaatecttcaaatctaaagcttcaaaacagcaaaacaattattccattaatgtttct
 55 ccatttctttctaaaatattaaaccaatataaaaaatcaaaagataacaattttgactaat
 acaaaaggatattttgcacttactgaagatggtaaagtgcacttgtgcttaaaacgcca
 cgtaataatgaaaaacatatgaaaatgccactgaatccactaaacctaagatttaaat
 gattttaaacaaactgcacagtaacaaaaccaactttagaatatcaaagtaacacgaaat
 gaaatgtacgcagaatatgtaaatcaattaaagaatttcagaatacagagaaacacaaggg

- tataatagttggtgtgccggctataccatgtcagcactattcaatgccacataataaca
aatcgatataatgcagaatcagtaatgagatatttacatcctaatttaagaggtcacgac
ttccaatttacaggactaacatctaacgagatgcttcgttttggtagatcacaaggcaga
aatactcaatatcttaatagaatgacttcatataatgaagtagaccaattaacaactaat
5 aatcaaggtatagctgtattaggttaagcgtgttgatcaagcgatggtattcacgctgga
catgccatggctgtggctggtaatgctaaagttaacaacggacaaaaagtcattttaatt
tggaaacccatgggacaatggctcatgactcaagatgcacatagtaatatcattccagta
tcaaatggcgatcactatgaatggatgcatcaatttatggttattaa
- 10 Sequence 668
MKKKLSYMITIMLAFTLSLALGLFFNSAHADSLPQKNGANQKTTKVTVSNKDVPDAVRKL
AEEQYLSRVALLDKASNHKATSYTLGEPFKIYKFNKESDGNYYYYPVLNKKGDVVYVVTIS
PNPSNSKASKQQNNYSINVSPFLSKILNQYKNQKITILTNTKGYFALTEGKVTLVLTKTP
RNNEKTYENATESTKPKDLNDFKQTASVTKPTLEYQSTRNEMYAEYVNLKNFRIRETQG
15 YNSWCAGYTMSALFNATYNTNRYNAESVMRYLHPNLRGHDFQFTGLTSNEMLRFGRSQGR
NTQYLNRMTSYNEVDQLTTNNQGI AVLGRVSSDGIHAGHAMAVAGNAKVNNNGQKVILI
WNPWDNGLMTQDAHSNII PVSNGDHYEWYASIYGY*
- Sequence 669
20 Contig_0502_pos_15222_15602,
putative peptide of unknown function
gtgatttcactactggttgtttgttattttttctgttggttcaccttcgccaactttt
tcccctgttaattgggttcttagttgttggtgttgtaattgtttttgttcctgggttcacct
ttctgtttaaacacgctcttcacctggtttaaatcaggattgaattcacgtttcttgtcg
25 aatggaatttctccggttgacgtaatcgaatctccatcaactggaccatattttgtcaca
tcatccactgggtgtgactacttcgcctgtatcaggatttttaactcctgggttacct
ggaacgtcctcttggtacctttcgggtgcgtttggatcaaattcatccttatggcctggc
ttgatttcttcgccaccataa
- 30 Sequence 670
VISSTGCFVIFSVGSPSPFTFSPVNGFLVVGVVIVFVPGSPFCLTRSSPGFKSGLNSRFLS
NGISSVDVIESPSTGPYFVTSSTGCVTTSPVSGFLTPLPGTSSWLPFGAFGSNSSLWPG
LISSPP*
- 35 Sequence 671
Contig_0502_pos_11871_10762,
is similar to (with p-value 2.0e-93)
>sp:sp|P53555|BIOA_BACSU ADENOSYLMETHIONINE-8-AMINO-7-OXONON
ANOATE AMINOTRANSFERASE (EC 2.6.1.62) (7,8-DIAMINO-PELARGONI
40 C ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE). >gp:gp|AF0
08220|AF008220_74 Bacillus subtilis rrnB-dnaB genomic region
. NID: g2293135. >gp:gp|U51868|BSU51868_2 Bacillus subtilis
biotin biosynthetic operon genes, complete and partial cds.
NID: g1277024. >gp:gp|Z99119|BSUB0016_96 Bacillus subtilis c
45 omplete genome (section 16 of 21): from 2997771 to 3213410.
NID: g2635411.
atgctattaggttcgtctaataattccatcgattgagctagccgaacagttagtc aaatta
acaccagatagattacaaaaagtgttttactctgatacaggagtgctcggtagagatt
gctattaagatggcttatcaatattggaagaatcgcgatgctgaacgatatgcgaagaag
50 aataaaatttcttacattacatcatggatcatcatggagatacaatagggttctgttagcgtt
gggtggtatcgatagtttccacaaaatttttaagaccttatttttgaaaatatacagata
gaaacaccgtgtttatataaaaagtaagtagcgaatgaagcggaatgcttaattcaata
ctgaatcaaatgaaaatatattatctgagagaaatgatgaaatagtaggatttattcta
gagccactatatacaaggtgcaacaggtttattcggttcacgcgatggtttttgaaagct
55 gtgaacagttatgtagaaaatagatgtattactaattttgtgatgaagtagcggttggt
ttcggacgtacgggagaaaatgtttgcttgtaaccatgaagatgtacaaccagatatattg
tgtctgggttaaggcgattacaggtggttatttaccgttagcggcaactttaacatctcaa
aagatatatgatgcttttttaagtcagagtcacggtaagaatacgtttttccacggtcat
acatatacaggtaatcagtttagttgttccgtagcacttgagaatattaatcttttttaa

aagaagcatctgattgggcacattcaaaagacatctcaaacattaaagcaacgcttagag
gcacttcaacctcataaaaatattggagatattagagggcggggattaatgtatggtgtg
gaattagttgaaaacaaatcaacgcagacaccactcgatattccaactgtagaactgatt
atacatcgatgtaaagagaatggattgatgattcgttaatttggaatatgtcatcactttc
5 gtacctattttaagtatgtctaataaagaattaaaaaaatggttaaaattttcaacaaa
gccttacatcaaacattgggtaagaagtaa

Sequence 672

MLLGSSNIPSIELAEQLVKLTDPRLQKVIFYSDTGSASVEIAIKMAYQYWKNI DAERYAKK
10 NKFLTLHHGYHGDITIGSVSVGGIDSFHKIFKDLIFENIQIETPCLYKSKYRNEAEMLSI
LNQIENILSERNDEIVGFILEPLIQGATGLFVHPHGFLKAVEQLCRKYDVLLICDEVAVG
FGRTGEMFACNHEDVQPDIMCLGKAITGGYLPAAITLSQKIYDAFLSQSHGKNTFFHGH
TYTGNQLVCSVALENINLFKKKHLIGHIQKTSQTLKORLEALQPHKNIGDIRGRGLMYGV
ELVENKSTQTPLDIPTVELIIHRCKENGLMIRNLENVITFVPILSMSNKEIKKMVKIFNK
15 ALHQTGKK*

Sequence 673

Contig_0502_pos_9610_8939,
is similar to (with p-value 2.0e-30)
20 >sp:sp|P53559|BIOW_BACSU_6-CARBOXYHEXANOATE--COA LIGASE (EC
6.2.1.14) (PIMELOYL-COA SYNTHASE). >gp:gp|U51868|BSU51868_1
Bacillus subtilis biotin biosynthetic operon genes, complete
and partial cds. NID: g1277024.
atgCGTgcaagccacgaagatattcatattagtggTgctgaaacaatgtgtgaatttgag
25 gatttagaaaaattatttaaaaaaatattttaataaagcatttaaatcatgaaaatggaat
atagatttctttaaatttgaaaattgaaaagggttaaggcaccgattcaaacggttagtagca
ttaccagtggTgaaatctaaacgatactttaacacaatttagcaaaacaaacaggtggt
tctgaatatgcgctaacaagggttagaatttataaaaaatgatattacttatactgga
gccattattctatctgcacaaaccggacaacgacttgatagcactgaacaacgaggtatc
30 agggtaacacaatttagcatttaaaacatgcaaatgtaattggagaaatatcagaaagagta
aaagatgcacgtgcacttgcaacttgatcaatgcatttgagggtgtaaaggcagaacta
tgtgtatcagacgatttgcatcacgactggatatttgctgcgcctaagtttaggatat
cgtagaattctttaaatttaaaagaaagggtacgcgtcacggagggaagaattatcttcgta
gacgaagaaataaattttaaatgaatatgtttccttttttagaaaacagtacctaagaaatc
35 atagaaaaataa

Sequence 674

MRASHEDIHISGAETMCEFEDLENYLKKYFNKAFNHENGNI DFLNLKIEKVKAPIQTLVA
LPVVENLNDTLTQLAKQTVSEYALNKGLEFIKNDITYTGAILLSAQTGQRLDSTEQRGI
40 RVTQLAFKTCNGEISERVKDARALATCINAFEGVKAELCVSDDLHYTTGYFASPKLGY
RRIFNIKEKGRHGGRIIFVDEEINLNEYVSFLETVPKEIIEK*

Sequence 675

Contig_0502_pos_6828_4888,
is similar to (with p-value 0.0e+00)
45 >gp:gp|AF090142|AF090142_1 Staphylococcus epidermidis lipase
precursor (gehD) gene, complete cds. NID: g3789931.
gtgatttttttgaaaaataataatgaaacaagaagatttagcatttaggaagtacacgggtg
ggagtcgtgtcaatcattactgggattacaatatttgTcagtggtcagcatgctcaagct
50 gctgaaatgacacaatcatcatcagattttaacgaacagtcacaacaaacagaacaagtt
gaacacaaagaagatacaactcatttatcatacgaattgaatcaagagggtgacacagct
agccaatcaagactaatcaagagaaccaatctgatgaaaatgtacaaaaaaagaataat
caaactcaacaagattcaacacaaacgtcaccattaaatgaccaagaacaaactttaag
gggcaacaatcaaaagacaatcatgttaccacaaattcacgtcaggatacatatccaaaa
55 ggccaaatcaagatgataaaggcaacaaacagtttaaaagataatcaacactcacaaaca
gaacatcaacctaatctcaaaacaaaataatgatcaagattcatcagataaaaaagcaa
cacccatctgatcaaaactcaagcccatcttcaaaaggaacacacctaacaacacag
tctataggagatagagataaaacagtaaaacaacatcttctaaagtacacaaaataggt
aatacaaaaactgataaaacagttaaaacaaatcaaaaaaagcaaacatcattaacttca

ccacgcgttgtgaaatcaaaacaaactaaacatatcaatcaacttactgcgcaagctcaa
 tataaaaatcaatatccagtcgtgtttgtacatggattttaggttttagtcggtgaagat
 tcattcagcatgtacccaattattggggtggtactaaatataacgtgaaacaagaactt
 5 acaaaattaggttaccgagttcacgaagccaatgtaggagcatttagcagcaattatgac
 cgtgctgttgaaactgtattattatattaaaggtggaagagtagattatggtgcagcacat
 gctgcaaaatattggtcacaagcgttatggcagaacatatgaaggcatcatgcctgattgg
 gaaccaggtaaaaaagatacatcttgttgacatagtaggtggccaaacgatacgcttg
 atggaacattttttaagaaatggaatcaagaagaaatagattatcaacgtcaatatggt
 ggtacggtatctgatttgtttaaggtggacaagataacatggtgtctacgattactaca
 10 ttaggaacacctcataatggcacacctgctgcagataaaactagggctgactaaatttatc
 aaagatacaattaatagaattggaaaaattggtggaactaaagcgctcgatttagaacta
 ggtttttctcaatgggcttcaaacagaaacctaatgaatcatatgctgaatatgcaaaa
 cgtatagcgaatagtaaaagtgtgggagactgaagatcaggctgtaaatgatttaacaact
 gttggagcagaaaagttaaaccaatgacgacattgaatcctaataatcgctctatacatca
 15 tatacagggtgctgcaacacatactggaccattaggcaatgaagtgccgaatattagacaa
 tccccactattcgatttaacaagtcgtgtgataggtggagatgataataaaaatgtcaga
 gtaaatgatggcatagtagctgtgtcttcttctactacatccaagtgtgaagcatttaag
 aaggtaggatgatgaacctagcaactgacaaggtatttggcaagtgaagcccgtaaaa
 tatgattgggatcatctagatttagtcggcttagatactactgattataagcgaactgga
 20 gaagaattaggtcaattctatatgagtagatgataaataatatgttgaaagtcgaagagtta
 gatggtattacacgtaagtag

Sequence 676

VIFLKNNETRRFSIRKYTVGVVSIITGITIFVSGQHAQAEMTQSSSDFNEQSQQTEQV
 25 EHKEPTTHLSYELNQEEDTASQSKTNQENQSDENVQKKNQTQQDSTQTSPLNDQEQTLK
 GQQSKDNHVTNSRQDTYPKGQNDKQKQFQDNQHSQTEHQPNQTNQNDQDSSDKKQ
 HPSDQTQAPSSKGTQPKQSISIGDRDKTVKQPSKVKHIGNTKTDKTVKTNQKKQTSLS
 PRVVKSKQTKHINQLTAQAQYKNQYPVVVFVHGFVGLVGEDSFSMPNYWGGTKYNVKQEL
 TKLGYRVHEANVGAFFSSNYDRAVELYYYIKGGRVDYGAHAHAQYGHKRYGRTYEGIMPDW
 30 EPGKKIHLVGHSMGGQTIIRLMEHFLRNGNQEIDYQRQYGGTVSDLFKGGQDNMVSTITT
 LGTPHNGTAAADKLSTKFIKDTINRIGKIGGTALDLELGFQWGFQKPNESYAEYAK
 RIANSKVWETEDQAVNDLTTVGAELKNQMTTLNPNIVYTSYTGAATHGTPLGNEVPNIRQ
 FPLFDLTSRVIGGDDKNVRVNDGIVPVSSSLHPSDEAFKKVGMNMLATDKGIWQVRPVQ
 YDWDHLDLVGLDITDYKRTGEELGQFYMSMINNMLKVEELDGITRK*

Sequence 677

Contig_0502_pos_4703_3888,
 putative peptide of unknown function

atgtatacaataatagagagatgtgaaaagatgaaatattatgggaagtgcatttcttac
 40 ataagcattttaataataacgttttttattggcggatgtggatttatgaataaagaaaat
 aataaagaagcggaaattaaagaaaattttaataaaacattaagtatgtatccaattaaa
 aatttagaagattttatcgataaaagaaggctatcgtagatgaagaatttgaaaaagaggac
 aaaggggacatggattattaattcagaaatgaatattcagaaaaaagatcaagcgatgaaa
 tctagaggtaggttttgtatatgaatagaataactagaaagacgactggtcatttttat
 45 acaaatataattacagaagataaaaaaggagagtgacacagtaaaagataaagaatatccg
 gttcgccctaaaaaacaataaaattgaaccgactaaacctatcgccgatgaaaaattaaaa
 aatgaattaaaaactttcagtttttctcctaataatgggaattttaaaatttaaaagac
 tacaagaatggaatgtgtcttataacccaaacgtaccaagctatttcggcagagtaccaa
 ttaagtaaatggaatgacaatgtgaagcaactcagaaagaggtatgatattccgattaaag
 50 agagctcctaaactaatattaaaaggggacgggtgaccttaaggttcattctataggttat
 aaagatatcgagttttcttttgcgacaataaagaagaagcgtctactttgcggatagt
 ttggaatttaatacgaagtgaagtaataatgagtag

Sequence 678

MYTIERCEKMKYYGKCISYISILILTFFIGCGFMKNKENKEAEIKENFNKTLSPYPIK
 55 NLEDLYDKEGYRDEEFKEDKGTWIINSEMNIQKKDQAMKSRGMVLYMNRNTRKTTGHFY
 TNIITEDKKGRVHSDKEYPVRLKNNKIEPTKPIADEKLKNEIKNFQFSQYGNFKNLKD
 YKNGNVSYNPNVPSYAEYQLSNEDDNVQLRKRYDPIKRAPKLILKGDGDLKGSSIGY
 KDIEFSFVDNKEESVYFADSLEFNPSEVNNE*

Sequence 679

Contig_0503_pos_4433_4828,

is similar to (with p-value 6.0e-30)

- 5 >sp:sp|Q02499|KPYK_BACST PYRUVATE KINASE (EC 2.7.1.40) (PK).
 >pir:pir|S29783|S29783 pyruvate kinase (EC 2.7.1.40) (version 2) - Bacillus stearothermophilus >gp:gp|D13095|BACPK_3 B. stearothermophilus phosphofructokinase and pyruvate kinase genes. NID: g285620.
- 10 atgccttttttctaaacctataattgcgcttggtgaagtaataaccattttcttctgtaatt
 agacctatagctttttcaacatatggtactaatgtttcatcaacagaatttgtaataata
 actttatcagataaatctttaccttctaaatcactagcactatctgcgacaattgcatgg
 cctacaacagatcctctaccaacaccttggccttttagcaatctcatcacctactaagtgg
 attttcacatatttgtagttcctttttctccagtaggtacaccagcagtaataataatt
 15 aaatctccgtttgaaactctaccagtttctactgctgttgctacagcattatttagtaaa
 gcacagttgttttacgtccttctttaacgacctga

Sequence 680

MPFSKPIIALGEVIPFSSVIRPIAFSTYGTNVSSTEFVIITLSDKSLPSKSLALSATIAW
 20 PTTDPLPTPWPLAISSPTKWIFIIFVVPFSPVGTAPAVIIKSPFETLPVSTAVATALFSK
 ASVVLRLPSLTT*

Sequence 681

Contig_0503_pos_4881_5186,

is similar to (with p-value 4.0e-31)

- 25 >sp:sp|P43659|SMPB_ENTFA SMALL PROTEIN B HOMOLOG. >gp:gp|M90
 060|STRATPASEA_1 Streptococcus faecalis H+ ATPase a (atpB),b
 (atpF),c (atpE),alpha (atpA), beta (atpD), gamma (atpG),delta
 a (atpH),and epsilon (atpC) subunits, complete cds. NID: g15
 3565.
- 30 gtgagacgagggcgaatgtacctgaataatgcatattgcaccatatgaagaagggaac
 cgttttaatcatgaccctttacgtacacgtaaaactcttgcacaaaaagaaattcaa
 aaattaggtgagcgtagacgagaaataggttattctattattccgttgaaagttaattta
 aaacatgggtcaatgtaaagttttattagcggttgctagaggtaaaaagaaatacgacaaa
 35 ctcaagcacttaaaagaaaagcggtaaaacgagatattgatcgcgagttaaagcccg
 tatta

Sequence 682

VRGEMYLNMMHIAPYEEGNRFNHDPLRTRKLLHKKKEIQKLGERTREIGYSIIPKLKYL
 40 KHGQCKVLLGVARGKKKYDKRQALKEKAVKRDIDRAVKARY*

Sequence 683

Contig_0503_pos_2851_1928,

putative peptide of unknown function

- 45 atgaagcctaaagtattgttagcaggtggcaccggctatattggtagatatttaagtcga
 gtcattgaacatgatgctcaattatgtgtttatctaaatatccaaacctgacaaagga
 tctacgaacaaaatcacatggttaaaacgcgatataataatcataaagatgtagttgaa
 gctatgaaaggtatagacattgcggtatattatttagatccaactaaacattctgctaaa
 ttaacacatgcaacagcagagatttaaaactttatagcggcagataaactttggtagagct
 50 gcatcaataaataagctgaaaaagattgtgtatattccggggagccgcatgataatgaa
 gctattgaacgtttaggcgcttatggcgtaccagttgattgtacggatgttgaaagtga
 cgctctcatattaacgtagaattacaaacagctaaatgatgatgttcgaacagcgatg
 aagatgattttaccaaagaaatggacgctcaatcaactttagactatttttagtaggtgg
 ttagatgagacaaaaggaactttgtacatactcaaaaacaagatcatcactacatcatt
 55 tacaattaggaacataaaagaccttttagctattttcaaaatgggttaatacaacagaagat
 ataattacattacatcttggtaggtgaaattgataaaactaaatcaagaagcaagca
 aaattagaatttagacttcttaaggaacacggttaattatgggttcatttatacgattat
 atccctagattattttggccaatttactattttcatacaagcaccgattcaaggacttctt
 atgagagggtttgaaattgattgtagaattaagcattatcaaggctcgattcaatcagg

gagaagattaaatataactaaataa

Sequence 684

5 MKPKVLLAGGTGYIGRYSRVIEHDAQLFALSKYKPKDKGSTNKITWLKRDIYNHKDVVE
AMKGIDIAVYYLDPTKHS AKLTHATARDLNFIAADNFGRAASINKLKKIVYIPGSRHDNE
AIERLGAYGVPVDCTDVEVKRPHINVELQTAKYDDVRTAMKMILPKKWTNLNQLVDYFSRW
LDETKGTFVHTQKQDHHYIIYNRNIKRLAIFKMVNTTEDIITLHLVDGKLMKPKSKKQA
KLEFRLLKGTRLIMVHLYDYIPRLFWPIYYFIQAPIQGLLMRGFEIDCRIKHYQGRIQSG
EKIKYTK*

Sequence 685

10 Contig_0503_pos_1660_1235,
is similar to (with p-value 3.0e-21)
>sp:sp|P42421|YXDJ_BACSU HYPOTHETICAL 26.6 KD SENSORY TRANSD
15 UCTION PROTEIN IN IDH 3'REGION. >gp:gp|D14399|BACIOLO_11 Bac
illus subtilis 15 kb chromosome segment contains the *iol* ope
ron. NID: g709980. >gp:gp|Z99124|BSUB0021_70 Bacillus subtil
is complete genome (section 21 of 21): from 3999281 to 42148
14. NID: g2636442.
20 atggatcaagtgatgagtgatggaacttggtgcagatgattatatgcaaaaaccattttat
acaaacgtcttaattgctaagctacaagctatttatagacgcgtttatgaatttgaggtt
gaagaaaagagaacgttaagttggcaagacgctactgtggatttatcaaaagatagtatt
caaaaagatgataaaactatctttttgtctaaaacagagatgattatttttagagatgta
atcaataaacgtaatacaatcgtagacacgagacactctcattactgctttgtgggatgat
25 gaagcttttggttagtgataatactttaacagttaatgttaataagattaagaaaaaatta
tcagaaattgacatggatagtgcaattgaaaccaaagttggtaaaggatacttagctcat
gaataa

Sequence 686

30 MDQVMSMELGADDYMQPFYTNVLI AKLQAIYRRVYEFGVEEKRTL SWQDATVDLSKDSI
QKDDKTI FLSKTEMIILEMLINKRNQIVTRDTLITALWDDAEFVSDNTLT VNVNRLRKKL
SEIDMDSAIETKVKGKGYLAHE*

Sequence 687

35 Contig_0503_pos_831_202,
putative peptide of unknown function
atgaaattattgatagatcaagagaatgatgatcagcgtaagcgagcggttattatttgaa
tggtctcgtattaatgagatgtagataagcaattatatttaacaaggcgttgaacacat
catcgtgatatgtattttgattatatttcattaaagagaatggttatagatgaatacaa
40 gttactcgacatatcagtcaggcaaaaggataggttttgaaattagattttaaaagacgaa
caaaagggtttatacagatgttaaatggtgccgtatgatgattaggcaagttctatctaac
tctttgaaatatagtgataattctacaataaatttaagtgggtataacatagaaggacac
gttggttttaaaaattaaagactacggtcgtggaattagtaaaagagatttaccacgtata
tttgatagaggatttacttctacaacagaccgcaacgatactgcgtcttctggtatggga
45 ttataccttgtacaaagcgtgaaagaacaacttgggattgaagttaaagttgattcaata
gtggggaaaggaacaacgtttttatttcattttcccaacaaaatgaaatcattgagcgc
atgtctaaagtgacaagattgtcattttta

Sequence 688

50 MKLLIDQENDDQRKRALLFEWSRINEMLDKQLYLTRLETHHRDMYFDYISLKRMVIDEIQ
VTRHISQAKGIGFELDFKDEQKVYTDVKWCRMMIRQVLSNLSKYSDNSTINLSGYNIEGH
VVLKIKDYGRGISKRDLPRI FDRGFTSTTDRNDTASSGMGLYLVSQVKEQLGIEVKVDSI
VGKGTTFYFIFPQQNEIERMSKVTRLSF*

Sequence 689

55 Contig_0505_pos_3663_4214,
is similar to (with p-value 3.0e-56)
>gp:gp|AF012285|AF012285_31 Bacillus subtilis *mobA-nprE* gene
region. NID: g3282109. >gp:gp|Z99111|BSUB0008_128 Bacillus

subtilis complete genome (section 8 of 21): from 1394791 to 1603020. NID: g2633699.

atgataacaatgaaagatattataagagatggatcatccaacacttcgtgaaaaagcgaaa
gaattaagcttcccacttttctaacaatgataaagaaacattgcgcgcaatgcgtgaattt
5 ctaatcaatagtcaggatgaagaaaccgcaaaacgttatggtttacgttctggcgtaggt
ttagctgctccacaaattaatgaaccaaaacgtatgattgctgtctacttacctgatgat
ggaaacggtaaatcgatgattatatgctcgtaaatcctaaaataatgagttacagtgt
caagaagcttattttaccaactggcgaaggttgtctaaagtgttgatgaaaacatcccaggt
ttagtgcacatgcacatagagtcactattaaagctcaagatattgatggaaatgatgtt
10 aaattacgtctcaaaggctatcctgcaattgtatttcaacacgaaattgatcatctaaat
ggcattatgttttatgattatattgatgccaatgaacctctaaaaccacatgaagaggcc
gtagaagtctaa

Sequence 690

15 MITMKDIIRDGHPTLREKAKELSFPLSNNDKETLRAMREFLINSQDEETAKRYGLRSGVG
LAAPQINEPKRMIAVYLPDDGNGKSYDYMLVNPKIMSYSVQEAYLPTGEGCLSDENIPG
LVHRHHRVTIKAQDIDGNDVKLRKGYPAIVFQHEIDHLNGIMFYDYIDANEPLKPHEEA
VEV*

20 Sequence 691

Contig_0505_pos_5404_6018,

is similar to (with p-value 1.0e-66)

>sp:sp|Q45493|YKQC_BACSU HYPOTHETICAL 61.5 KD PROTEIN IN ADE
C-PDHA INTERGENIC REGION. >gp:gp|AF012285|AF012285_29 Bacill
25 us subtilis mobA-nprE gene region. NID: g3282109. >gp:gp|Z99
111|BSUB0008_125 Bacillus subtilis complete genome (section
8 of 21): from 1394791 to 1603020. NID: g2633699.

atggctcaattaggtcatgaaggtgtgttgcttactatcagactcaacaaacgcacta
gttccagattttactttaagtgaacgtgaagttggacagaatgtcgataaaattttcaga
aattgtaagggcgtaggtatctttgcaacttttgcttctaataatttatcggtgttcagcaa
30 gcagttgaagcagcaattaaatataatcgtaaaatcgttacatttggacggttcaatggaa
aacaatatcaaaattgggtatggaactaggatataatcaaagcgccaccagaaacgtttata
gaacctataaaaataaatagtgtaacaaacacaggttactcattcttctgactggttct
caaggtgaacctatggctgcattatcaagaattgcaaatggtacacataagcaaatataaa
35 attataccggaagacactgtagtatttagttcttcgcctattccaggttaacactaagagt
atcaatcgtaacattaatgcgttgtaacaaagctggtgcagatgtgattcatagtaaaatt
tcaaacatcttgcaaattttagatgctacttcggttccccatatacaagctttttcaat
actttggttatctag

40 Sequence 692

MAQLGHEGVLCCLSDSTNALVPDFTLSEREVGQNVDKIFRNCKGRIIFATFASNIYRVQQ
AVEAAIKYNRKIVTFGRSMENNIKIGMELGYIKAPPETFIEPNKINSVPKHELLILCTGS
QGEPMALSRANGTHKQIKIIPEDTVVFSSSPIPGNTKSINRTINALYKAGADVIHSKI
SNILHNFRCYFVSPYTSFFNTLVI*

45

Sequence 693

Contig_0505_pos_3483_2857,

putative peptide of unknown function

atgaatttttaaaaagactgtagcaattgtcctaacgtcagcagtggtatttagctggatgt
50 actatagataaaaaagaaattaaaaaatatgatgatcaagtacaaaaagctatggaccaa
gagaaaaccgttaataagtaagtaaaaaataaacgaattagaagagaaaaagcaaaaa
ttattttaaaaggtaaatgataaagatcaaagcacacgtaaaaaagcagctgaagatata
gttgaaaatgtaaaacaaagacaaaaagaatttgaaaaagaagagaaggctctagataat
tctgaaaaagcatttaacaaagccaagcaaatatcttgaacatgtagaaaacaaagcaaa
55 aaaaaagagttgaaacacttgatagtgctatttaagaaaaataaatcacatgatgct
tatgcaaaggcttacaaaaagcacttaataagaaaaagaactgttttcttatttgaat
gaagataatgcaacacaatcggaagtagacggaatcgaaagatctttctaaagcatat
aaagaaatgaataataaatttaagtcttactcaaaagccattgagaaagtaaaaaagagaa
aaacaagatgtagaccaattaaaaataa

Sequence 694

MNFKKTVAIVLTSVLLAGCTIDKKEIKKYDDQVQKAMDQEKTVNQVSKKINELEEKQKQK
 LFKKVNDKDQSTRKKAEDIVENVKQRQKEFEKEEKALDNSEKAFKQAKQYLEHVENKAK
 5 KKEVEQLDSAIKEKYKSHDAYAKAYKKALNKEKELFSYLNEDNATQSEVDGKSKDLSKAY
 KEMNNKFNAYSKAIEKVKREKQDQVDQLK*

Sequence 695

Contig_0505_pos_2686_1574,
 10 is similar to (with p-value 0.0e+00)
 >pir:pir|S10798|DEBSPF pyruvate dehydrogenase (lipoamide) (E
 C 1.2.4.1) alpha chain - Bacillus stearothermophilus >gp:gp|
 X53560|BSPDMC_3 B. stearothermophilus pdhA, pdhB, pdhC, pdhD
 genes for pyruvate dehydrogenase multienzyme complex (E.C.
 15 numbers 1.2.4.1, 2.3.1.12, 1.8.1.4). NID: g40038.
 atggctcctaagttacaagcccaattcgatgcagttaaagttttaaatgagactcaatcg
 aaatttgaaatgggttcaaattttggatgaagacggaaatgtcgttaatgaagacttagta
 cctgatttaacagacgaacaatttagtggaattaatggaaagaatggatggactagaatt
 cttgatcaacggttctatttcgttaaatagacaaggacgttttaggtttctatgcaccaaca
 20 gcaggacaagaagcttcacaattagcatctcagtatgcttttagaaagtgaagacttcatt
 ttacctgggttatcgtgatgtgcctcagattatttggcatggcttacctcttacagacgca
 ttcttattctcaagaggacacttcaaaggttaaccaattccctgagggagttaatgcactt
 agccctcaaattattatcgtgcacaatatattcaaactgccggtgtagcgtttggactt
 aaaaaacgtggcaaaaatgcagtcgcaattacttatacaggtgatgggtggttcacacaa
 25 ggtgacttctatgaaggaattaactttgcatctgcatacaaaagcacctgcaatttttgta
 attcaaaacaataactatgccatctctacaccacgtagtaaacacacagctgcagaaaca
 tttagcacaagaagctatttcagttgggtatccctggaattcaagttgatggatggatgct
 ttagctgtttatcaagcaacattagaagcacgtgaacgtgcagtagcaggagaaggtcct
 actgttatcgaaactttaacttatcgttatggaccacatactatggctgggtgatgatcct
 30 actcgttatagaacttcagatgaagatgctgaatgggagaaaaagaccatttagtacgt
 ttcagaaaaatatttagaagctaaaggtctttggaatgaagacaaagaaaatgaagtgggt
 gaacgtgcaaaaatctgaaataaaagcagctattaaagaggctgacaatacagaaaaacaa
 actgttacttctctaattggatatcatgtatgaagaaatgcctcaaaatttagcagaacaa
 tatgaaatttacaaagagaaggagtcgaagtaa

Sequence 696

MAPKLQAFQDAVKVLNETQSKFEMVQILDEDGNVNVNEDLVPDLTDEQLVELMERMVWTRI
 LDQRSISLNRQGRGLGFYAPTAGQEASQLASQYALESEDFILPGYRDVPQIIWHGLPLTDA
 FLFSRGHFKNQFPEGVNALSPOIIIGAQYIQTAGVAFGLKKRGKNAVAITYTGDGSSQ
 40 GDFYEGINFASAYKAPAI FVIQNNNYAISTPRSKQTAETLAQKAISVGI PGIQVDGMDA
 LAVYQATLEARERAVAGEGPTVIETLTTRYGPHTMAGDDPTRYRTSDEDAWEKKDPLVR
 FRKYLEAKGLWNEDKENEVVERAKSEIKAAIKEADNTEKQTVTSLMDIMYEEMPQNLAEQ
 YEIYKEKESK*

Sequence 697

Contig_0505_pos_1570_593,
 is similar to (with p-value 0.0e+00)
 >pir:pir|C36718|C36718 pyruvate dehydrogenase (lipoamide) (E
 C 1.2.4.1) E1 beta chain precursor - Bacillus subtilis >gp:g
 50 p|AF012285|AF012285_34 Bacillus subtilis mobA-nprE gene regi
 on. NID: g3282109. >gp:gp|M57435|BACPYDHY_3 B.subtilis pyruv
 ate dehydrogenase complex genes, complete cds; PAL-related l
 ipoprotein (slp) gene, complete cds, lysine decarboxylase (c
 ad) gene, partial cds. NID: g143375. >gp:gp|Z99111|BSUB0008_
 55 131 Bacillus subtilis complete genome (section 8 of 21): fro
 m 1394791 to 1603020. NID: g2633699.
 atggcacaaatgacaatggttcaagcgattaacgatgcgcttaaaagtgaactcaaaaga
 gacgaagacggttttagttttcggtgaagacggttggtgtaacggtgggtgtattccgtggt
 actgaaggtttacaaaagaatttggcgaagatcgagtatttgatacaccattagcagag

tctggaattggtgggcttgactaggttagcagtgactggcttccgtcctgttatggaa
 attcaattccttaggattcgtttatgaagtatttgacgaagtagctggtaaattgctcgt
 actcgtttccgttcaggtggaactaaaccagcgctgttacaattcgtacaccttttgg
 5 ggtggcgtccacactccagagttgcatgctgataatttagaaggtatcttagctcaatca
 cctggtttgaagtagttattccatcaggtccttatgatgctaaaggattattaatttct
 tctattcaaaagtaattgatccagttgtatatctagaacatatgaaattatatcgttcttc
 cgtgaagaggttcctgaagaagaatacaaaattgacattggaaaagccaatgttaaaaaa
 gaaggaatgatattactctaatacttacggggcaatggtacaagaatcactaaaagct
 gctgaagagttagaaaaagatggttattcagttgaagttattgacttacgtactgtacaa
 10 ccaattgatatagatacttttagtagcatcagttgagaaaactggacgtgctgtagttgta
 caagaagcacaacgtcaagctggtgtgggtgcacaagtggcagcagaattagcagaagcga
 gcaattccttcattagaagctccaatagctcgagtagccgcacagatacaatttatcca
 ttactcaagctgaaaacgtttggttaccaaataaaaaagatattatagagcaagctaag
 gcaactttagaattctaa

15 Sequence 698
 MAQMTMVQAINDAKSELKRDEDVLVFGEDVGVNGGVFRVTEGLQKEFGEDRVFDTPLAE
 SGIGGLALGLAVTGFRPVMEIQFLGFVYEVFDEVAGQIARTRFRSGGTPAPVTIRTPFG
 GGVHTPELHADNLEGILAQSPGLKVVI PS GPYDAKGLLISSIQSNDFVVYLEHMKLYRSF
 20 REEVPEEYKIDIGKANVKKEGNDITLISYGAMVQESLKAAEELEKDGYSVEVIDLRTVQ
 PIDIDTLVASVEKTGRAVVVQEAQRQAGVGAQVAELAERAILSLEAPIARVAASDTIYP
 FTQAENVWLPNKKDIIIEQAKATLEF*

Sequence 699
 25 Contig_0505_pos_462_151,
 is similar to (with p-value 8.0e-44)
 >sp:sp|Q59821|ODP2_STAAU DIHYDROLIPOAMIDE ACETYLTRANSFERASE
 COMPONENT (E2) OF PYRUVATE DEHYDROGENASE COMPLEX (EC 2.3.1.1
 2). >pir:pir|S19722|S19722 dihydrolipoamide S-acetyltransfer
 30 ase (EC 2.3.1.12) chain E2 - Staphylococcus aureus >gp:gp|X5
 8434|SAPDHDNA_2 S.aureus pdhB, pdhC and pdhD genes for pyruv
 ate decarboxylase, dihydrolipoamide acetyltransferase and di
 hydrolipoamide dehydrogenase. NID: g48871.
 gtggcatttgaatttagattaccgatatcggggaaggtatccacgaaggtgaaattggt
 35 aaatggtttattaaagccggcgatacaattgaagaagatgatgtattagcagaagttcaa
 aatgataaattctgtagtagaaattccttctccagtaagtgttactgttgaagaagtgtta
 gtagatgaaggaacagtgccagtagtaggagatgtcatcgttaaaattgatgcacctgat
 gcagaagaaatgcaatttaaaggtcatggcgatgatgaggattctaagaaagaagaaaaa
 40 gaaatgatttga

Sequence 700
 VAFEFRLPDIGEGIHGEIIVKWFIKAGDTIEEDDLAEVQNDKSVVEIPSPVSGTVEEVL
 VDEGTAVVVDVIVKIDAPDAEEMQFKGHGDDDEDSKKEEKEMI*

45 Sequence 701
 Contig_0506_pos_1522_2664,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q45493|YKQC_BACSU HYPOTHETICAL 61.5 KD PROTEIN IN ADE
 C-PDHA INTERGENIC REGION. >gp:gp|AF012285|AF012285_29 Bacill
 50 us subtilis mobA-nprE gene region. NID: g3282109. >gp:gp|Z99
 111|BSUB0008_125 Bacillus subtilis complete genome (section
 8 of 21): from 1394791 to 1603020. NID: g2633699.
 atggctcaattaggtcatgaaggtgtgtgtgtgttactatcagactcaacaaacgcacta
 gttccagattttactttaagtgaacgtgaagttggacagaatgtcgataaaattttcaga
 55 aattgttaagggcgctattatctttgcaacttttgcttctaataatttatcgtgttcagcaa
 gcagttgaagcagcaattaaatataatcgtaaaatcgttacattttggacgttcaattggaa
 aacaatatcaaaattggtatggaactaggatatataaagcgccaccagaaacgtttata
 gaacctataaaaaataatagtgtacctaaacacaggttactcattctttgtactggttct
 caaggtgaacctatggctgcattatcaagaattgcaaatggtacacataagcaaaataaaa

attataccggaagacactgtagtatttagttcttcgcctattccaggtaacactaagagt
 atcaatcgtacaattaatgcgttgtaaaaagctgggtgcagatgtgattcatagtaaaatt
 tcaaacattcacacttctggacacggttctcaaggatgaacaattaatgttacgtctg
 attcaacctaaatacttccctgccaatcacaggatgaatcgtatgcttaaaagctcatggt
 5 gagactgggtgttcaatgcggtgttgatgaagataatgtattttttcgatatcggtgat
 gtacttgctttaacacatgattctgcacgaaaagcaggaagaattccttccggcaatgta
 cttgttgatggcagtggtataggtgatattggcaatggtgtcatcagagatcgtaaat
 ctttcagaagaagggttagttattgttgttgagcattgactttaataactaacaatta
 ctatctggccctgatattatttcacgcggttttgtttatatgcgggaatctggtcaatta
 10 atttatgatgctcaacgtaaaattaaaggcagatgtcatttctaaacttaacagcaataaa
 gatattcaatggcatcaaatataatcttcaattatcgaaacattacatccttatctttat
 gaaaaaacagctcgaaaacctatgattttacctgtgataatgaaagtaaatgaagataaa
 taa

15 Sequence 702

MAQLGHEGVLCLLSDSTNALVPDFTLSEREVGNVDKIFRNCKGRIIFATFASNIYRVQQ
 AVEAAIKYNRKIVTFGRSMENNIKIGMELGYIKAPPETFIEPNKINSVPKHELLILCTGS
 QGEPMAALSRIANGTHKQIKIIPEDTVVFSSSPIPGNTKSINRTINALYKAGADVIHSKI
 SNIHTSGHGSQGDQQLMLRLIQPKYFLPIHGEYRMLKAHGETGVQCGVDEDNVFIDIGD
 20 VLALTHDSARKAGRIPSGNVLDGSGIGDIGNVVIRDRKLLSEGLVIVVVSIDFNTNKL
 LSGPDIISRGFVYMRESGQLIYDAQRKIKGDVISKLSNKNQIWHQIKSSIETLHPYLY
 EKTARKPMILPVMKVNEBK*

Sequence 703

25 Contig_0507_pos_639_1073,
 putative peptide of unknown function
 atgaagcaagcttttagaaaaatatttacaagcgaatagcgatgtacttgataataagtat
 gtcatgcaacataaattagataaaacaaagtgatagtaatcctaaaatcacagaatcacia
 gctgatcgtcttagcaagttatccaatttagcagttaagaacgatttacatttcaaaaaa
 30 ttataaaaaacaatcacatccctgaagaatataaagatccaacagatcgcataattaat
 tattttcacgctttaaatagtagcatttcaaagttagatgaagacattgagaaattaaac
 taccaaccacaaaattcaattaacgttgttgatgtagccacaaaatattcagggtgatgta
 aataaaaaacaacagataaaattactactttccttaagaaaaaaggaatagacacagaa
 gtatttaataataaa

35 Sequence 704

MKQALEKYLQANSOVLNKNYVMQHKLDKQSDSNPKITESQADRLSKLSNLAVKNDLHFKK
 FIKNNHIPEEYKDPDRIINYFHALNSTISNVDEIDIEKLNYPQNSINVVDVATKYSQDV
 NKKQQDKITTFLLKKKGIDTEVFNK*

40 Sequence 705

Contig_0507_pos_1278_1670,
 putative peptide of unknown function
 atgaggaaaatcattatgaagatacgtttaacatttattatcttagcaatactatccacc
 45 atcggttagtacttggttttagcaaaaatccaacaggccacacacaaatcaactataac
 gaaccttatacagtactcatagccattacgacaatagttataatggctttaccagcactc
 atattaggtatatttaatcatcttgcatgtagaatcatatcggcgatattacaaataagt
 gcactgatgatgtgggggttttttagtaaatcattagcttaattatgggacaaattgtcatt
 atgcttatggcttcttaacgatacttgcatcttcttagttctattgtcacactttca
 50 gtgcacccatctacttcagataaaataaattaa

Sequence 706

MRKIIMKIRLTFIILAILSTIGLVVLAKYPTGPHTINYNEPYTVLIAITIVIMALPAL
 ILGIFNHLACRIISAILQISALMMWGLVLIISLIMGQIVIMLMASLTILALLVSSIVTSL
 55 VHPSTSDKIN*

Sequence 707

Contig_0507_pos_3386_2754,
 is similar to (with p-value 2.0e-17)

>gp:gp|U93874|BSU93874_16 *Bacillus subtilis* cysteine synthase (yrhA), cystathionine gamma-lyase (yrhB), YrhC (yrhC), YrhD (yrhD), formate dehydrogenase chain A (yrhE), YrhF (yrhF), formate dehydrogenase (yrhG), YrhH (yrhH), regulatory protein (yrhI), cytochrome P450 102 (yrhJ), YrhK (yrhK), hypothetical protein YrhL (yrhL), putative anti-SigV factor (yrhM), RNA polymerase sigma factor SigV (sigV) and YrhO (yrhO) genes, complete cds, and YrhP (yrhP) gene, partial cds. NID: gl934604.

10 atggacgggtttaattacattttatcattatcacattattgattattatagtagctgggcca
gatttcattattgtaatgaaaaacactattaattcaagtaaaatgaatggttttatggct
gcatttgggtattactacggggcatattttatactcttcattagctatttttgaattata
tatactactacgagtttacactttgttttttaacaataaaaatattgggtgcttgttat
cttattttatctcggaatcaaaagtatttttgagtgcgcacagttctgttgatttttagtaaa
15 caagcttttagctgatgtcagaaatgtgagttatatcacttcttttagacaaggttttta
agtacaagctttaaccctaaggctttattttttatgtagtatattcccgcagttcctt
tctaacggtaatacatatgaaatctgaagttgcgttatttgcttttcagttgttgta
gttatatgcttatggttttttttgtgtattcatcttcaatatattaaattattattc
agcagaccgagattcaaggctatatttgattatattgtagggtttgttttaattggctta
20 tctattaatttattattaagtaaaagtagctaa

Sequence 708

MDGLITFIIITLLIIIVPGPDFIIVMKNTINSSKMNGFMAAFGITTGHILYSSLAIFGII
YILTSLHFVFLTIKILGACYLIYLGIKSILSAHSSVDFSKQALADVNRVSYITSFRQGF
25 STSLNPKALLFYVSIFFQFLSNGNIHMKSEVALFAFSVVVVICLWFLFCVFIFQYIKLLF
SRPRFKAIFDYIVGFVLIGLSINLLLSKSS*

Sequence 709

Contig_0508_pos_3124_1019,
putative peptide of unknown function

30 atgacatgttttttaatatataaagaatatcattctccccctgaacgcaatattaat
agaattattctgatgttttcacttacaaatgacttaaaaaatcactattaatgggtgaaaca
aaagatttaggtaatcatatagctatcattaatcaatctgacatctattttattaatagt
gcttcaaatctcgattactctctattccagttattttttatagtaaaagataataaaa
35 ttttttaaatgttattttgacagacattttattacaatcaagcagttttgttaaaacaatt
attttacaagctattcaacatttaataaaaaggagaaaatcaagatgagcaatccatctct
aaaataatacaaacgctactaaaagaagcagtcattcgatataagaaaaaatatattcct
caaattgcagttaatcattcagtggttactgaaggattaacatttattcattcaaaaagta
tcacagtcactttcactacgagaagtagctcaacattgtaatatatctgaatcttattgt
40 tctaacctattttgcaagatatcttaatatgaatttttaagattattttacaagcttaaaa
gtgattgactctataaaaaggctactttcatctgaggactcaattaacgctatttcagaa
caatctggatttagtagtcataccaattttacaaatcaatttaaaaattatttaggtgt
agcccaaaaacaataaccgaacgattatctctaagtttagactccttaccttcgataagtttt
agtgatactgacttttcacaatatattgatttaattaatcaatttgagtttagtgatcat
45 ttggctactgaaacgactgaaagagatatcaatgaattttatcctcaagatcagactaaa
aactctaagcgtttatacgttttcaaaaatttcaacgaattatttcaatttgttttaat
gaatattacaacattgattttacctccctgccacaagctgtaatttttatcaatgatatac
actgatatttcgacgcgagaggtgaacttttaatttataaatcgatgctttgaaaaatta
ttcgaaaaaaatataggttttagccatgagattaacatctacaaatgaatttgaaatctatc
50 aaagaaataattttattattttcttaatatgccaccaagattataaaatgaacaaaaaaatg
gttaatttatgttagtctttgaaactaagaatatgtcagtgaaacgatatacatttatgc
catttaaaaattaaaaataaaaataaagctatccgttatagtataactgttgaggatta
ctacacaaaaattcttctattgatcgaaacttatgatatgatgaagcgattaaatttcgat
tactattttatagacattgaaaatctagaaacaaaaaactcattaattaccaaacgtaaa
55 tcgtacttacattcgctccacacactttgaaaattataaacaatttatattagattccggt
ataccttcaactaaatttggtttacaataatttatctttaaagtgttttaatatacaaac
aatggtagcttatccacttcaattatctgaccttggttgctatttagtcgcattaatgcgt
tacgggggtggtgtaagttatcaactgatagaagatgagagtccttttattgccttattt
aatcgttatggtagtcccctacctctcatgcacctctataaattaatcgaaccattttta

aatgaaccttttagagatagctaacaattttttaatgagtcgcaaagatggtaactatcac
 tttttattatttaataaaataaatgatcggttatctatctgatagtcacacgcgtacgtt
 tttaaaaatacattatcaaccaactcattaattattattaaaaacggttaaatcatgagcat
 5 ggcgcgaattcaaaaccttctaccacaaactaaacaacaatttttatattgaacgtagtatt
 cttgatgaacttgataaatcaaatcaacaaaaacagaatttagctatacaacatgaccat
 catcttcctttccaagtcaccttaaaacacgatgaagtcaaatatatttggttttaaacct
 tcttaa

Sequence 710

10 MTCFLNIYKEYHSPPERNINRIILMFSLTNDLKITINGETKDLGNHIAIINQSDIYFINS
 ASNLVLLSIPVIYFYSKDNKFFKCYFDRHLLQSSSFVKTIILQAIQHLLIKGENQDEQSI
 KIIQTLLKEAVIRYKKKIYIPQIAVNHSVFTEGLTFIHSKVSQSLSLREVAQHNCISESYC
 SNLFARYLNMNFKDYFTSLKVIDSIKRLLSSEDSINAISEQSGFSSTHTNFTNQFKNYLGC
 SPKQYRTIISKLDLPSISFSDTDFSQYIDLINQFEFSOHLATETTERDINEFYPPQDQTK
 15 NSKAFIRFQNFELFQFVFNEYNYIDFTSLPQAVIFINDITDISTREVNFNLLNRCFEKL
 FEKNIGLAMRLTSTNEFESIKEIILLFLNSHQDYKMNKMKVFMVLFETKNMSVNDIHL
 HLKIKNKNKAIRYSITVEGLLHQNSSIDRTYDMMKRLNFDYFYFIDIENLETKNSLITKRK
 SYLHSSSTHFENYKQFILDGIPSTKFVYNNLSLKCFKYTNNGTYPLQLSDLVCHLVALMR
 YGGVSYQLIEDESPFIALFNRYGSPPLMHLYKLIEPFLNEPLEIANNFLMSRKDGNH
 20 FLLFNKINDRYLSDSQORYVFNKNTLSTNSLIIKTLNHEHGAIQNLLPQTKQQFYIERSI
 LDELKSNQPKTELAIQHDHHLPFQVTLKHDEVKYICFKPS*

Sequence 711

Contig_0508_pos_988_524,
 25 is similar to (with p-value 2.0e-25)
 >sp:sp|P80238|GS26_BACSU GENERAL STRESS PROTEIN 26. >gp:gp|A
 B001488|AB001488_7 Bacillus subtilis genome sequence, 148 kb
 sequence of the region between 35 and 47 degree. NID: g1881
 226. >gp:gp|Z99106|BSUB0003_69 Bacillus subtilis complete ge
 30 nome (section 3 of 21): from 402751 to 611850. NID: g2632653

atgtgtgataaggatatatacaataataaggagatgataattttgaataaacaacaagtg
 acaaaaagcaatagaaaaagattaaattcttcaaaaattgggtgcctatcaacagcacat
 cataataaacctaatagcagatatatgggtcttttacaacgatgacttaaaccttatataca
 35 aaaacgaatatcaattcactaaaagtgcgaagaatagaaaataatcctgatgctcatatt
 ttattaggctataacgaaacaacaacaatagctttgttgaaatagatgccactatagaa
 gttgtcaaaaatcaaaaagttattgattgggttgggaaactcaagacaaaacatttttc
 aattcaaaaagaagatcctgaattatgtgtactcaaagttatacctcggttcaattaaatta
 atgaatgatgatgaactagatacgccagctacaattgagttataa

Sequence 712

MCDKDIYNNKEMIILNKQVTKAIEKVLNSSKIGVLSTAHHNKPNSTRYMVFYNDLNL
 45 KTNINSLKVEEIEENPDAILLGYNETTNNSEFVIDATIEVKNQKVIDWLWETQDKTFF
 NSKEDPELCVLKVIIPRSIKLMNDELDTPATIEL*

Sequence 713

Contig_0509_pos_308_706,
 putative peptide of unknown function
 atgatagaagaattgattaaccgtgaaaaaatgaattttgggtgaatcaatatactttta
 50 cagtttgggtatgttaaaagaagatatgaagttgccaaaatcttatatttttgaaattgct
 tccaactggaagaaaattggattttcaaatgcaaaacagcatatgaatatgcattacaa
 gttaatcaacctaaaaattacgaaacacattctaataataacgacagacaatcggtgga
 agacaaaatcaatttttatccaaagaaaagacacctaaatggcttcaaaaataggacgat
 caagaagaaaataaagaataaatgatgacactctcgaagaagatcgacaagcatttctt
 55 gaaaagttaaatcaaaagtgaaggaggaagataactaa

Sequence 714

MIEELINREKMNFVINILLQFVMLKEDMKLPKSYIFEIASNWKIGISNAKQAYEYALQ
 VNQPKNYETHSNDKRQNNRGRQNQFLSKEKTPKWLQNRRDQEENKEINDDTLEEDRQAF

EKLNQKWKEEDN*

Sequence 715

Contig_0509_pos_727_1194,

5 putative peptide of unknown function

atgggcgattctcaaaatctagataaacgtatacaaaaaataaaacaaaatgtaatcaat
gatactgacgttaaaccattttcttgagaaaaatcgtagtaataataactaatgagatgata
gacgaagatttaaatgttcttcaagagtataaagatcaacaaaaagtttatgatggacat
cgctatgatgattgtccgaattttgtaaaaggacatgttcctgaactatatattgaaaat
10 gaaagaatcaaaattagatatctaccttgcccgtgtaaaattaaacatgatgaggaacga
tttgattcacaacttattacatctcaccatattgcaaagagatacacttcatgcaaagctc
aaagatatttatatgaataatcgagagagacttgatgtagcaatggcagctgatcaaatc
tgtacagcaattactcaaattaaaattagaaagttttatacgtcctaa

15 Sequence 716

MGDSQNLDKRIQKIKQNVINDTDVKHFLEKNRSNITNEMIDEDLNLVQKEYKDQQKVYDGH
RYDDCPNFVKGHVPELYIENERIKIRYLPCPKIKHDEERFDSQLITSHHMQRDTLHAKL
KDIYMNNRERLDVAMAADQICTAITQIKIRKFYTS*

20 Sequence 717

Contig_0509_pos_1283_2650,

is similar to (with p-value 3.0e-20)

>sp:sp|P10564|HEXA_STRPN DNA MISMATCH REPAIR PROTEIN HEXA. >
pir:pir|C28667|C28667 DNA mismatch repair protein hexA - Str
25 eptococcus pneumoniae >gp:gp|M18729|STRHEXA_3 S.pneumoniae m
ismatch repair protein (hexA) gene, complete cds. NID: g1536
54.

30

atggatacgttattttcataaaaattaatttttaatttcactgcaattgggtgaaatgcgacta
tatgcaactttaagaggtatgtttaaggtaaatcaaacctcattgataaacatgtttaaa
gaaaaataaagtatttcgtttaaatgtatcttacattctttctaaaattggtaaaaatgta
taccctttgtttccagatcaaatgttaccactaaagcgaaatattttattaatgtttgt
ccgttggtaccatttatcggttcgcattcatttttttaattccttcaaaaggatatatta
atatgtcttacttttatgatttttaaatgcaatatttatctttcaactaaaaaatcttat
gaccaagatttaaaatcaattttttatactgctaattgttataaaagcaaagtcaagcttta
35 agtaagattgagagcacgcccgcgataagtgttgattttactcattttaaagcttcacgc
cgttttagtggtttattagctagagtagaatcacaagatatggcgagtagcataatcatg
tttattaaattagtagtattcatgatagattatgttttatttcatttaatacaacgcagctac
tttaagtatcaagaagaagttagacatgttatgactacataagcatattagataatcat
tactctatagctatgtatcaacatactttgacacattattgttatcctaaaatcaatcac
aatattgaatggtcttcaaatgaaatcaatcattcatcctctactagatgaagaaaatgcg
attgctaacacaattgacatttcaaatcatatattgctcacaggctctaattgcatcagga
aaatctacatttatgaaagcagttgcactaaatttgatttttagctcaatcgatacaaaact
gcaacagctcactcattttatcaacctggctatgtaattgacatcaatggcaaattgcg
gatgacgttttaagtgtgacagttatttcatgtcagaacttaagtctattcgtagatta
45 ttttaacactcatcagtgaataagatatattgttttatagatgaaatttttaaaagaacg
aatacaactgaacgtattgcggtctctgaatcagtattatcgtatttagataatcaaaaa
gcatatcaggttatcgctgcgacacatgatgttgaattatcaacattattagaaaataca
tataataattatcatttttaaatgaatcaattcaagaaaatagcatatttttcgattacaaa
attaaaccaggtaagccaatacacgtaattgaattactacgcattacgcagttt
50 cctatcgatattttatcagcgtgctcaacaaaatttcgaaacctctag

Sequence 718

MDTLFHKINFNFTAIGEMRLYATLRGMFKVNQTSLINMFKENKVFRNLNVSYILSKIGKNV
YPLFPDQMLPTKRNIILMFCPLLPFIGFAFIFLIPSKGILICLTFMILNAILSFKLKKS
55 DQDLKSIFYTANVIKQSQUALSKIESTPAISVDFTHFKASRRFSGLLARVESQDMASSIIM
FIKLVFIMIDYVLFHLIQRSYFKYQEEVMTCDYISILDNHYSIAMYQHTLTHYCPKINH
NINGLQMKSIHPLLDENAIANTIDISNHILLTGSNASGKSTFMKAVALNLILAQSIQT
ATAHSFIYQPGYVMTSMANADDVLSGDSYFMSSELKSIRRLFNTHQCNKIYCFIDEIFKRT
NTTERIAASESVLSYLDNQKAYQVIAATHOVELSTLLENTYNNYHFNESIQENSIFFDYK

IKPGKANTRNAIELLRITQFPIDIYQRAQQNIRNL*

Sequence 719

Contig_0509_pos_9141_8665,

5 putative peptide of unknown function

atgctacgttattctctacgcacagcatcgacacagtcacaaattagcagaagaattaagc
tacattgagcagtgatgttgccatacaaaaatccgcttcgatgatatgatacagctttac
atcgatgctactgaggggtgtacaacatcaaacaattggtaagatgatgcttcaaccactc
gtagaaaatgccatcaagcatggctcgtgatagtgaaacctttaagataacaattcgtatc
10 agacttacgaagcgcaaaattacataattctggttcgatgataatggcatcggtatgtctcca
tcacatttagaacagtgcgccaatcccttcacacgatgtttttgatacgacacaccta
ggtttaaatcattttacataaatagagccatgattcaatatggaacatatgcacgtctgcac
atcttctcaagaagccagcaagggacattaatgtgttaccataataccacttgtctag

15 Sequence 720

MLRYSRLTASHTVKLAELSIEQYVAIQNIRFDDMIQLYIDATEGVQHQTIGKMMLOPL
VENAIKHGRDSEPLKITIRIRLTKRKLHILVHDNGIGMSPSHLEHVRQSLHHDVDFDTHL
GLNHLHNRAMIQYGYARLHIFSRSQGTLMCYQIPLV*

20 Sequence 721

Contig_0509_pos_8653_7898,

putative peptide of unknown function

atgtttaaagtagttatttgtgatgatgaaaggattataagagaaggcttaaagcaaagt
gttccatgggaggactatcatttcaccactgtttatactgccaaagacggcgtggaagca
25 ttgtctttaattcgccaacatcaacctgaactcgctcattactgatatacgaatgcctcga
aaaaatgggtgttgacctactagatgacatcaaagaccttgattgccagattatcatttta
tcgagttatgacgacttcgaatatatgaaagccggtatacaacatcatgttcttgattat
ttactaaagccagtagaccacactcagttagagcatattctagacatattagttcaaagg
ttattagaacgcccacattctaccaatgatgacgcggcatatcatactgcctttcaacca
30 ttattaaaaaattgattacgatgactattatgtcaatcaaatTTTTgtctcaaatcaagcaa
cattatcacaagaaagtgactgttcttgacttaattaatcctattgatgtaagtgaagtca
tacgcttaggagcgttttaagaacatgtaggcattacgatagttgattatctaaatcgt
tatcgatttttaaaatcattacatcttttagaccagcactacaagcattatgaaattgct
gaaaaagtaggtttttctgagtataaaatgttttgctatcattttaaaaaatatttacat
35 atgtcaccaagtgattataataagcaatcaaaatag

Sequence 722

MFKVVICDDERIIREGLKQMPWEDYHFTTVYTAKDGEALSLIRQHQPVLITDIRMPR
KNGVDLLDDIKDLDCQIIILSSYDDFEYMKAGIQHHVLDYLLKPVDHTQLEHILDILVQR
40 LLERPHSTNDAAAYHTAFQPLLKIDYDDYVQNILSQIKQHYHKKVTVLDLINPIDVSES
YAMRTFEKHVGITIVDYLNRIRILKSLHLLDQHYKHYEIAEKVGFSEYKMFYHFKKYLH
MSPSDYNKQSK*

Sequence 723

45 Contig_0509_pos_6213_5707,

is similar to (with p-value 5.0e-25)

>sp:sp|P43984|Y318_HAEIN HYPOTHETICAL PROTEIN HI0318. >pir:p
ir|B64006|B64006 hypothetical protein HI0318 - Haemophilus i
nfluenzae (strain Rd KW20) >gp:gp|U32717|U32717_5 Haemophilu
s influenzae Rd section 32 of 163 of the complete genome. NI
50 D: g1573283.

atggaggacatcatgattttaactattttatttatctttttctgtattcgactcatcagc
ttaagatatctatgcaacacgcaaaacagctaaaggtagagggcgcggtggaatatggt
gtgaaaaattcaaaatatctagccattacgcatgtattaatttacatgagtgacgtata
55 gaagcattcattcgtaaggatacatttagtctacttaacggcattggcttaatcatattg
atcatcgcttatatcatgctatttatagttattaagacattaggtcgtatttggaattg
aaattattttactgcccacacccattattattaagtacgggttatataaagtaacgaaa
catccaaactattttttaaatatcattcccgaatttaattgggtgtattactactaacaat
gctacatacacacactcttatttagttccatagcttattttttaattgtacgtatccgt

caagaagagaaattaatgaatatataa

Sequence 724

MEDIMILTILFIFFCIRLISLISKISMQHAKQLKVEGAVEYGVKNSKYLAITHVLIYMSAAI
5 EAFIRKDTFSLNIGLIIILIIAYIMLFIVIKTLGRIWTLKLFILPNHPIIKSGLYKVTK
HPNYFLNIPELIGVLLLLTNATYTTLLLVPIYAYFLIVRIRQEEKLMNI*

Sequence 725

Contig_0509_pos_5264_4449,
10 is similar to (with p-value 2.0e-68)
>gp:gp|U30714|BAU30714_2 Bacillus anthracis Weybridge A toxin
plasmid pX01 right inverted repeat element (WeyAR) borderi
ng the toxin-encoding region, ORFA and ORFB genes, complete
cds. NID: g929970. >gp:gp|U30715|BAU30715_2 Bacillus anthrac
15 is Sterne toxin plasmid pX01 left inverted repeat element (S
terneL) bordering the toxin-encoding region, ORFB and trunca
ted ORFA genes, complete cds. NID: g929973.
gtggatcaattaaaagtaaaatattcaatcaaattgatactagaagtattaaacatacct
aatcaacatattaccgatggaaaaacaaaacctataaaaatgataccgtaacacaaaaa
20 gtcattgaattatgtaaagctaaccactatacctacgggttatcgtaagattacagcattg
attaatcaatgtttacatcaccaattaatcataagagagtagacagagaatgatgcagaag
catcatttgaactgccgaggttagacctaaaaagacgacaagaataggtaaaccgtattat
aaaacggacaattttattacaaagacaatttaaagcgagttgtcccatggaagtattaaca
accgatattactttattaccatttggctcattctatgtttgtattttatcttcgataatggat
25 atttataacggagaaattgtggcgtataaaaatagatgataaacaagaccaaagtttagtt
aatgatacatattaaatcaaatcgatatacctgaaggttgatatattacatagtgatcaaggc
agcgtttatacatcttatgcttattatcaatttgtcgaagaaaaaggcattatcagaagt
atgtcccgaaggaacacctgccgataacgccccgatagaaagtttccattcctcgcta
aagtctgaaactttttacatcaataatgagcttaatcgctctaattcatattgtaatagat
30 attgtcgaaaagtacattaaaaactataataataatcgaattcaacaaaaactaggctac
ttatccccagtaaaaatacagagaattaatagcctag

Sequence 726

VDQLKVKYSIKLILEVLNIPKSTYYRWKNKTYKNDTVTQKVIELCKANHYYTYGYRKITAL
35 INQCYTSPINHKRVQRMQKHHLNCRVRPKKTTRIGKPYKTDNLLQRFKASCPMEVLT
TDITYLPFGHSMYLLSSIMDIYNGEIVAYKIDDKQDQSLVNDTLNQIDIPEGCILHSDQG
SVYTSYAYYQLCEEKGIIRSMRKGTPADNAPIESFHSSLKSETFYINNELNRSNHIVID
IVEKYIKNYYNNRIQQKLGYLSPVKYRELIA*

40 Sequence 727

Contig_0509_pos_3554_2733,
is similar to (with p-value 2.0e-47)
>gp:gp|AL031317|SC6G4_30 Streptomyces coelicolor cosmid 6G4.
NID: g3449234.
45 gtgagtgcaaaagtgaagaatggaagacattgacgctattgcagtaacacaaggcccagga
ttaataggagctttatttgattggtattaatgcggctaaagctttggcatttgcttatgat
aagcctattattccagtagatcatattgctggtcatatttatgccaatcacttagaacia
ccattaacgtttccactaatgtcattgattgtatctggtggtcactgaactagtatat
atgaaaaatcatttagatttgcgaagtgttggtaaacgagagatgatgcagtaggagaa
50 gcttatgataaagttgctcgaacaatcaatcttccttatcctggtggaccgcatattgat
cgattagcagctaaaggtaaagatgtatatgattttccaagagtttggtctgaaaaagat
agttatgatttttagtttttagtggtcttaaaagtgtgtgaataaataaaactgcataattta
agacagaaaaatattgaattgtagctgaagatgttgcaacgagtttccaaaatagtgtt
gtagaagttttaacctataaagctattcatgcttgtaaaacttataatgttaatcgctta
55 attgttgtaggtggtggtgcttagtaataaaggattaaagaaatgcactaagtgaagcatgt
aaaaaagagggtatacaccttactattccaagtcctgttctttgactgataatgcagcg
atgattggtgctgctggatattatttatatcaagctggtttgctggtgagatttagcttta
aatggacaaaaataattgatattgaaactttttctgtttaa

Sequence 728

VSAKVKMEDIDAIIVTQGPGLIGALLIGINAAKALAFAYDKPIIPVHHIAGHIYANHLEQ
 PLTFPLMSLIVSGGHTLVYMKNHLDVEVIGETRDDAVGEAYDKVARTINLPYPGGPHID
 RLAAGKGDVYDFPRVWLEKDSYDFSFSGLKSAVINKLHNLQKNIEIVAEDVATSEFQNSV
 5 VEVLTYKAIHACKTYNVNRLIVAGGVASNKGLRNALSEACKKEGIHLTIPSPVLCTDNAA
 MIGAAGYYLYQAGLRGDLALNGQNNIDIETFSV*

Sequence 729

Contig_0510_pos_315_650,
 10 is similar to (with p-value 4.0e-19)
 >gp:gp|L42945|STALYTS_1 Staphylococcus aureus lytS and lytR
 genes, complete cds. NID: gl854576.
 gtggaaaattctattaaacatgcatttaaaatcgtaaaaagaataatcatattgatgtg
 gatgttagcatgaagcaagactacttaagtatatctgttcaagataatggcaggcata
 15 ccagctgatcaattagatactattggatatacgacagtaacgtctaccactggactgg
 aatgccttagtcaatcttaataaaaagacttactggactatttggacaacatcggcactg
 aacattcaatcttctcaatcaggcagcactgtaagttgtttaattccatataaatcttct
 aaggaggaacactttaatgaaagcgtaaatcggtga

Sequence 730

VENSIKHAFAKNRKNNHIDVDVSMKQDYLSISVQDNGQGIPADQLDTIGYTTVTSTTGTG
 NALVNLNKRLTGLFGTTSALNIQSSQSGTTVSLIPYKSSKEEHFNESVNR*

Sequence 731

Contig_0510_pos_799_1389,
 25 is similar to (with p-value 2.0e-53)
 >gp:gp|L42945|STALYTS_2 Staphylococcus aureus lytS and lytR
 genes, complete cds. NID: gl854576.
 atggatgaaagtggattgatttagctcaaaaaattaataaaatgaagcgatcaccacat
 30 attatctttgcaaccgctcacgagaaatttgcagtcaaagcctttgaattaaatgaacc
 gattatatattaaaaccttttgaaaaagaacgtattaatcaagctgtaaatagggttgac
 atggctaaagataaatcaaaaaacaaagataaaaactatcacacctaataatattgattat
 agtcatgatgagcgctcaaacacatgtactcccaattgaagtggatgaacgtattcac
 atcttaaaatttcacagacattatcgcatattatctgttaataatgggattacaacgatagat
 35 acaacaaaacaaagtattgaaacgaccgaaacacttaatcattacgagaaaaaactacct
 tcctctctattttattaaaatacatcgcgctactatcggttaataaagaacatatccaaaca
 atagagcattggtttaattatacgtatcagctgacgttaacacatgaatttaatatcaa
 gtttagtcgttcttatatgaagacttttaacaacaacttggtcttcaataa

Sequence 732

MDESGIDLAQKINKMRSPHIIFATAHEKFAVKAFELNATDYILKPFKERINQAVNKVD
 MAKDKSKNKDKTITPKYIDYSDDERAQTHVLPVIEVDRIHILNFTDIIALSVNNGITTID
 TTKQSYETTETLNHYEKKLPSSLFIKIHRAITVKNKEHIQTIEHWFNYTYQLTLTHEFKYQ
 VRSYMKTFKQQLGLQ*

45

Sequence 733

Contig_0510_pos_1584_1982,
 is similar to (with p-value 3.0e-28)
 >gp:gp|U52961|SAU52961_1 Staphylococcus aureus holin-like pr
 50 otein LrgA (lrgA) and LrgB (lrgB) genes, complete cds. NID:
 gl841516.
 atgaaatacagtatTTTTTcaacaagcattaacgattgcagtgattttacttatatcaaaa
 attattgaatcatTTTatgcctattccaatgccagcttcagtaattggacttgactatta
 tttatcgcatTTgtgtacaggcattgtgaaattaggtcaagttgagactgtgggaactgca
 55 ttaaccaataatattggattcctatttcgtaccagccggtatttcagtcattaaactcttta
 ccaatccttaagcaaagccctattTTtaattttttacttattatttttcaaaccttttta
 ttattaattttgactggcttttgcgtcacaattatttagtgacgaaatcacttttcccttct
 aaagagaaaaatgaagaacaagtcacataggagggttaa

Sequence 734

MKYSIFQQALTIIVILLISKIIESFMPIPMASVIGLVLLFIALCTGIVKLGQVETVGTA
LTNNIGFLFVPAGISVINSLPILKQSPILIIILLIIISTLLLLICTGFASQLLVTKSLFPS
KEKNEETSHIGG*

5

Sequence 735

Contig_0510_pos_1986_2687,

is similar to (with p-value 2.0e-80)

>gp:gp|U52961|SAU52961_2 Staphylococcus aureus holin-like pr
10 otein LrgA (lrgA) and LrgB (lrgB) genes, complete cds. NID:
g1841516.

atgattgaacatttaggaattaatacaccttattttgggatattagatcattaataacca
tttgtcatagcgacttattttataaaaaaacgaatggttctttttactagcaccttta
ttcgtgaagtattggtgcaggtattgctttttgaaattgacaggaattagttatgagaat
15 tataaaatcgggtggcgacattattaatttctttctagaaccagctacaatatgctttgcg
attccttttatatcgcaagcggaagtattaaaaaggtattggttacaataatttgggtggt
atagctggttggtacaattattgccttggttattaatttattcttgggtgcaataacattccaa
tttggcaatcaaattatagcatctatgctacctcaagctgcaacgacagcaattgcatta
cctgtatctgacggtatcgggtggtgtcaaagaattaacctcactcgcagttattttaaat
20 gcagttgtcatttctgcttttaggtgctaaaatagttaaattattttaaaatatctaaccct
attgccagaggacttgactagggaagaagtggaacacacttttaggtgtcgcggcagctaaa
gaattgggtgagactgaagaatcaatgggaagtattgcagttgtcatcggtggcggttatt
gttgtagcagtagttcctatccttgctccaatcttattataa

Sequence 736

MIEHLGINTPYFGLVSLIPFVIATYFYKKTNGFFLLAPLFVSMVAGIAFLKLTGISYEN
YKIGGDIINFFLEPATICFAIPLYRKREVLKRYWLQIFGGIAVGTIIALLLIYLVAITFQ
FGNQIIASMLPQAATTAIALPVSDGIGGVKELTSLAVILNAVVISALGAKIVKLFKISNP
IARGLALGTSGHTLGVAALKELGETEESMGSIADVIVGVIVVAVVPILAPILL*

30

Sequence 737

Contig_0510_pos_3536_3931,

is similar to (with p-value 1.0e-34)

>gp:gp|AF009352|AF009352_4 Bacillus subtilis osmoprotectant
35 transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
(opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388.

atgattgaatgtaacaaattacctttcattacttatgaccacctttcttttcttcaaagt
aatgatgtttcttttaaatattcttcagctatcactgcaggttccttaccttttccatccg
40 cttcataatttaacttctgcatcttctctgttgagattttaccttctaatttttttagtg
ccttatcgatttctggattatcctttattaattgttcatttgcaagtggaactaccgtcat
aaggcgggaagaatttgcgatcatcttccaatattttcaaataagctgcaatacgtc
catctgttgaatacccaactgctacatctaattttatttttttaatgcatcatatacta
45 aaccaatttgcattggacgtgcactatcaaatttaa

Sequence 738

MIECNKLPFITDYHLSFLQSNVSLNILQLSLQVPYLFHPLHNLTSFLLLRFYLLIFLV
PYRFLDYPLLIIVHLQVDYRHKAGRICDHLPIFSNHKLQYVHLLNTQLLHLIYYFLMHHL
50 NQFALDVHYQI*

Sequence 739

Contig_0510_pos_6909_0,

is similar to (with p-value 7.0e-53)

>gp:gp|L35343|PSEACOX_5 Pseudomonas putida TPP-dependent ace
55 toin dehydrogenase alpha and beta-subunits (acoA and acoB),
dihydrolipoamide acetyltransferase (acoC), g2,3-butanediol d
ehydrogenase (adh) and acoX genes, complete cds. NID: g52955
9.

atgaaagcagcagtatggtatggacaaaaggatgtacgcgttgaagatcgcgaaacccaaa
gcaataaaagacaatgaagtgaagttaaagtctcttgggcccgtatctgtggtactgat
ttacatgaatatttgaaggacctatctttatttcaactgatcaaccggacccactactt
5 ggtcaaaactgcacctgtgacttttaggtcatgaattttcaggtgtcatagaaaatggtggt
aaagacgtatcacgttttaaaaaaggggatcgtgtggttagttaatccaacagtgctctaaa
agagaaaagccggaaaatggtgacttgtatgatggttattcatttataggactaggttct
gatggtgcatttgcgagtttactaatgctcctgaaacaaatggttatcatctaccagat
aatgtttcagcagcagaaggtgctcttgtagaaccaacagccgttgctgtccaagcagtt
aaagaaggcgaattattatttcggtgatactgtagcagttttggcgctgggccaattggt
10 ttgttaactattgttgcagcaaaagctgctggtgcaagtaaaatatttgcctttgactta
tcagaagaacgttttagcgaaaagctaaaagtgtcgggtgcgactcacgtgtataactcaggt
aacgtcgatccagtaaaacggtttatgaacacactgacaacggtgtagatgtgtcattt
gaagttgctggtgtaggattactttacaacaatctattgaagtaaacacgtccgcgtggt
actgctgtcatcgatcaatcttcggtcatcccgtagaattcaatccattattacaaatg
15 aataaagggtgtcaagttaacaactacaattgc

Sequence 740

MKAAVWYGQKDVREDREPKAIKDNEVQVKVSWAGICGTDLHEYLEGPIFISTDQPDPLL
GQTAPVTLGHEFSGVIENVGKDVSRFKKGDRVVVNPTVSKREKPEENVLDYDYSFIGLGS
20 DGAFAEFTNAPETNVYHLPDNVSAREGALVEPTAVAVQAVKEGELLFGDTPAVFGAGPIG
LLTIVAAKAAGASKIFVFDLSEERLAKAKSVGATHVYNSGNVDPVQTVYEHTDNGVDVSF
EVAGVGITLQOSIEVTRPRGTAVIVSIFGHPVEFNPLLQMNKGVKLTTTIA

Sequence 741

Contig_0510_pos_6273_5167,
is similar to (with p-value 2.0e-94)
>gp:gp|AF009352|AF009352_2 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
30 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_69 Bacillus subtili
s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: g2635827.
atgattgaggcgacagatggacagattatgatgaatggaaaagatgtccgtaatatgaat
35 cctgttgaattgcggagaagatcggttatgtcattcaacaaattggtttgatgccacat
atgactattcgagaaaatattgttttagtacctaaacttttaaaatggtctaaagagaag
aaagatgaaaaagctaaagaacttattaaactggttagatttacctgaagaatatttggat
cgttatccagctgaattgtcaggagggaacaacaacgaattggtgtgtgcgcgcttta
gcagctgaacaagatattatattaatggatgaacctttcgggtgcattagatcctattaca
40 cgcgatacattacaagatttagtaaaaggaattacaacaaaaattaggaaaaacatttatt
tttgtcactcatgatattggtgaggctattaaattagcagacaaaatattgtattatgtct
aagggaaaagtcgttcaatacgtacatacactgacaatattttacgatatcctgcaaatgac
tttgttagagattttattgggcaaaatcgcttgattcaggatcgctcctaataatgaaatct
gtggaaagtgtatgatcaaaccgctcactgttaaagcagatgattcattgaatgatgca
45 gtaaatattatgagaacacgtcgagtagacactattttttagtcaataatcaaaataaa
ttattaggatttttagatattgaagatatcaatcaaggattacgtgcgcgtaaaagaatta
attgataccatgcgaaagggatgtctacaaagtacatatcaattcaaagttacaagactca
gtgcgtactatttctaaaacgtaattgttagaaatgtccctgtggtcgataatgatgaacat
ctcatttggttttaattacacgtgcaaaacttagtcgatattgtgtatgactcaatttggggc
50 gaagaagattctgatagttatgagatcccaaatgaaagcttagatgagaataatcacgat
ttaccacaaaatcaaactgatacacgaacaaatataaatgaagatgtgaatgattatcat
gatgctcaacatagaggtgaggattaa

Sequence 742

MIEATDGQIMMNGKDVNRNMPVELRRSIGYVIQQIGLMPHMTIRENIVLVPKLLKWSKEK
55 KDEKAKELIKLVDPPEEYLDYPAELSGGQQQRIGVVRLAAEQDIILMDEPFGALDPIT
RDTLQDLVKELQQKLGKTFIFVTHDMDEAIKLADKICIMSKGKVVQYDTPDNILRYPAND
FVRDFIGQNRLIQDRPNMKSVEAMIKPVTVKADDSLNDVNMRTTRRVDITFVVNNQNK
LLGFLDIEDINQGLRARKELIDTMQRDVYKVHINSKLQDSVRTILKRNVRNVPVVDNDEH

LIGLITRANLVDIVYDSIWGEEDSDSYEIPNESLDENNHDLPQNQTDTRTNINEDVNDYH
DAQHRGED*

Sequence 743

5 Contig_0510_pos_5164_4529,
is similar to (with p-value 5.0e-55)
>gp:gp|AF009352|AF009352_3 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
10 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_68 Bacillus subtili
s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: g2635827.
atgaaagcattcttacaagaatattggtagtagtcaacttttatcaaaagcagtagaacatttt
15 tatatttctatgtttgcattattgtagcgattgtagtagctgtcccttttaggtatttta
ttatcaaaaacgcaacgcacagctaattgtggtattacagttgctggcgtgcttcaaacc
attcctactttggctgtgctagctatcatgattccaatatttgggtaggaaaaacacca
gctattgttgcaattatttatctatgtattattaccaattttaataatacagtagtaggt
gttaaaaaatatcgataaaaaatgtcattcaagctggcacaagtagggaatgactaaattt
20 caattaatgaaagatgtagaaatgccttttagctttaccacttattattagtggtattcgt
ctatcaagtgtagtatacgtcatttagttgggcaacactcgcaagttatgtaggtgcaggtgga
cttggggatcttgatttaattggattaaatctctatcaaccacctatgattattagtgca
gcgattgttggtactttattagcatttagttattgactttatactttcatttagttgaaaaa
tggtgtgtacctaaggattaaaagtagtatctagataa

Sequence 744

MKAFLQEYGSQLLSKAVEHFYISMFALLLAIVVAVPLGILLSKTQRTANVVLTVAGVLQT
IPTLAVLAIMIPIFGVGKTPAIVALFIYVLLPILNNTVLGVKNIDKNVIQAGQSMGMTKF
QLMKDVEMLALPLIISGIRLSSVYVISWATLASVYGAGGLGDLVFNGLNLYQPPMIISA
30 AIVVTLALVIDFILSLVEKWVVPKGLKVS*

Sequence 745

Contig_0510_pos_4461_3568,
is similar to (with p-value 7.0e-85)
35 >gp:gp|AF009352|AF009352_4 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
(opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388.
40 gtgttatctggatgcagtttaccaggttttaggtgatggaaatgcaaaagatgatgtgaaa
atcacacagactgaaacaagtgaactaagattataggtcatatggaaaaattattaatt
gaacatgaaactgatggaaaaatcaaaccgacgttgattgggaacctaggttctagcatt
attcaacataatgcgtttacaacgtggtgatgcaaatatgtcagcggtacgttacacaggt
actgaattgacgagtgatttagcagctaaacctactaaagatcctgataaggccatgtct
45 gaaacacacgcttattttaaaagaaatatgatgaaaagtattatcattcacttggttt
gcgaatacatatcgattcatggtgacaaaagaaacggctaaaaagtagtacttagaaaaa
gtatcagatttagagaaatataaagatgaactacgtcttggaatggatacccaatggatg
aaccgtgcaggtgatggatatccagcttttgtaaagattatggatttaaatttgatagt
gcacgtccaatgcaattggttttagtatatgatgcattaaaaaataataaattagatgta
50 gcagttgggtattcaacagatggacgtattgcagcttatgatttgaaaatattggaagat
gatcgcaattcttcccgcttatgacggttagtccacttgcaaatgaacaattaataaag
gataatccagaaatcgataaggcactaaaaaaattagaaggtaaaatctcaacagaagaa
atgcagaagttaaattatgaagcggatggaaaaggtaaggaacctgcagtgatagctgaa
gaatatttaaagaaacatcattactttgaagaaaagaaagggtggtcataagtaa

Sequence 746

VLSGCSLPLGDLGNKDDVKITTTTETSETKIIGHMEKLLIEHETDGKIKPTLIGNLGSSI
IQHNALQRGDANMSAVRYTGELTSVLAAPTKDPPDKAMSETQRLFKKKYDEKYYHSLGF
ANTYAFMVTKETAKKYHLEKVSLEKYKDELRLGMDTQWMNRAGDGYPAFVKDYGFKFDS

bmr) and its regulator (bmrR) genes, complete cds, and branched-chain 2-oxo acid dehydrogenase (bfmB) gene, 3' end. NID: g2558636. >gp:gp|D84432|BACJH642_251 Bacillus subtilis DNA, 283 Kb region containing skin element. NID: g2627063. >gp:gp|299116|BSUB0013_111 Bacillus subtilis complete genome (section 13 of 21): from 2395261 to 2613730. NID: g2634723.

5 atgtgtaaacacctctctcttcaactcagtgaaaataaaaggcgatattattaaatattgt
 aaatctattaaaaatgaaaattatagctctgatgtagacggttttattttatttaggtgga
 gatggtacacttaataactagtaaatggcggttatgcagtatcagttaaattttaccaatc
 10 ggtgtaataaccaggtggttacctttaacgattttacaaaaacacttcaactgcacccta
 tttaaaacagctagtgagcaattattaacatcacatgctgaatcatatgatgtgttaaaa
 gtgaacgacttatatgtacttaatttcggttgacttggttaatagtacaaaatgcagag
 aatgttcaagatggttctaaagatatattcggttaaatcagctatattggatcaaccgtt
 aaaacggttattaaatcctgtttaaatttgatttctcattgactggtgatggtgaaacaaa
 15 gaaggcaataacttcgatgatgttaataagcaaacggtcccaatataggtggtggacaaatt
 ccgctaaccgatttatcgccacaagatggaagagcaaacacatttgtatttaaatgatcaa
 acactaaatataattgaatgatataataaaaaaacgtgatagtagaattggaacgaaatc
 acacaaggtattgatcacatatcaggttaagcacatcacactctcaacaaaacccatgatg
 aaagtggatattgatggcgaaattaatttagaaacaccaattgagattcaagtattaccc
 20 aaagcgatacaacttcttactgcaactgaacaaaataattaa

Sequence 752

MCKHLSLQLENKGDIIKYCKSIKENYSSDVLFLILGGDGLNELVNGVMQYQLNLPI
 GVIPGGTFNDFTKTLQLHPNFKTASEQLLTSHAESYDVLKVNLDLYVLNFVGLGLIVQNAE
 25 NVQDGSKDIFGKFSYIGSTVKTLLNPVKFDFSLTVDETKEGNTSMMLIANGPNIGGGQI
 PLTDLSPQDGRANTFVFNDDTLNLDILKKRDSMNWNEITQIDHISGKHITLSTNPSM
 KVDIDGEINLETFIEIQVLPKAIQLLTATEQNN*

Sequence 753

30 Contig_0511_pos_3293_642,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q24803|ADH2_ENTHI ALCOHOL DEHYDROGENASE 2 (EC 1.1.1.1
) (ADH) / ALCETALDEHYDE DEHYDROGENASE (EC 1.2.1.10) (ACDH).
 >gp:gp|U04863|EHU04863_1 Entamoeba histolytica HML:IMSS alcoh
 35 ol dehydrogenase 2 (EhADH2) mRNA, complete cds. NID: g48842
 9.
 atgtttgtgaattatttcacaatatctaaggagtggtgttatatgttatctgtaactaaa
 aaaaatacatatgaatcaaacaaagatgaagtcacacaaatgattgattcattagcagaa
 aaaggacaagaagctctaaaaagaactatctaaaaaatcacaacatgagattaatgacatt
 40 gtacatcagatgagcatggctgctgtgatcagcatatgcatttagctaaactagcttac
 gcgaacacaggttagaggtattttatgaagacaaagctatcaaaaatttatatgcctcagag
 tacatatggaattcaatcaaagacaataaaacagtaggtatcataggtgaagataaaacaa
 aaaggattaacgtatgtagctgaacctataggcgtgatttgtggagtgacaccaacgacc
 aaccctacatctacaactattttcaaagcaatgattgctattaaaacaggtaatccaatt
 45 attttcgcatttcatccaagtgcacaacaatcatcaaaatatgctgctaaagtcatttta
 gaagctgcaacaaaagcaggtgctcctaaagattgtattcaatggatagaagtgccatca
 attgaggcaactaaacaattaatgaatcataaagatatgtctttagttctagcgactgga
 ggctctggaatggttaaagtcgcataatcgacaggtaaacctgcattaggagtcggtcca
 ggtaatgttctacttatattgaaaaaactgctcatatcaaacgtgctgttaatgatatc
 50 attggttctaaaacttttgataatggtatgatttgtgcttctgaacaagtcaggtgtt
 gataaagaagtatacactgacgtcgttaaagaattcaaatcaccaaacatattttgtt
 aataaaaatgaactacaacaattagaagatgccatcatgaatgaagataaaactgcagtt
 aaacctgatatagttggttaaattcgtctgtagatatagcgaaattgtcaggaattagtgtt
 ccagaaaaaacaaaattattagtcgcagaaattgatggaattggaagattatccttta
 55 ccagtgaaaaattatcacctgtactcgcaatggttaactgcaaaatcaacaggacatgca
 ctacaattttgtgaagacatatataaaatttggtggttaggtcacactgctgaattcac
 accgaggatagtcaattacaacaaaaattcggtctaaaaatgaaagcttgccgtgatttg
 gtaaatcaccttctgctgctcgagggaattggaatatgtataatgaactcattccttca
 ctcacgttaggttgtggttcatatggtagaattctatttctcataatgaagcgagta

gacttattaaatattaaaacaatagcaaaacgctcgtaataatatgcaatgggttaaactc
 ccacctaaagtttattttgaagaaaattcagttatgtatttgacagagatggataatggt
 gaacgtgtaatgatagtttgtgatccaggaatggtaatatgggttatactgatatagtt
 5 gaacaagtgctgagacgccgagaaaaaccaaccacaaatcaaagtgtttacgaagttgaa
 cctaattccatcaactcatcacgtctataaggggttagaaatgtttataaaattccaacct
 aataactattattgcactcgggtggcgggttcggcaatggatgcagccaaagcaatatggatg
 ttctttgagcatccagaaaacttcattttttggggcaaaacaaaagttcttagatattcgt
 aaacgtactttataaaattaccaaacctaaaaacgcaaaatttatatgtataccaacgaca
 10 tcaggaactgggttctgaagtgcacacctttgcagtaattactgatagcgagacacacgtt
 aagtatccactagcagattatgcgttaactcccgatattgctatcgatccacaattc
 gtattaagtgtacctaaagatggtgccgcagatacaggaatggatgttttgacacatgcc
 attgaatccttacgtctctgtcatggcttcagattatacaagaggcttaagcttacaagca
 ataaagttaaacttttgattatctaaaaatcatcagttcaagaaaatgacaaacactcacga
 15 gaaaaaatgcataatgcttcaacaatggccgggtatggcatttgccaatgcttttttagga
 atttctcattctatcgacataaaattgggtggtgaatatggtattccccacggcagaaca
 aatgctatttttattaccacatgtcattcgctataatgccaagatccacaaaaacatgca
 ctgtttcctaaatatgatttcttttagagcagatactgactatgctgacattgcaaaat
 ttaggactcaaaggttaatacaactgaagaattagtggtatgctctagctaatacggtgtat
 20 gatttaggatgttcagttgggtattgatatgaattttaaatacacaaggcgttaactgaagag
 cttcttcactctactatagacagaatggctgaattagcatttgaagatcaatgtacaact
 gctaataccaaaagaaccgctaattagtgaaacttaaaggcattatcgaaacagcatatgat
 tatgaagataa

Sequence 754

25 MFVNYFTISKWLYMLSVTKKNTYESNKDEVQIMIDSLAEKGQELKELSKKSQHEINDI
 VHQMMAAVDQHMHLAKLAYDETRGIYEDKAIKNLYASEYIWNISKDNKTVGIIGEDKQ
 KGLTYVAEPIGVICGVTPTTNPSTTIFKAMIAIKTGNPIIFAFHPSAQSSKYAAKVIL
 EAATKAGAPKDCIQWIEVPSIEATKQLMNHKDIALVLATGGSGMVKSAYSTGKPALGVGP
 30 GNVPTYIEKTAHIKRAVNDIIGSKTFDNGMICASEQVMVVDKEVYTDVVKEFKLHQTYFV
 NKNELQQLEDAIMNEDKTAVKPDIVGKSAVDIAKLSGISVPEKTKLLVAEIDGIGKDYPL
 SREKLSPVLAMVTAKSTGHALQICEDILKFGGLGHTAVIHTEDSQLQKFKGLMKACRVL
 VNTPSAVGGIGNMYNELIPSLTLGCGSYGRNSISHNVSVDLLNIKTIKRRNNMQWFKL
 PKVYFEENSVMYLTENDNVERVMI VCDPGMVNIGYTDIVEQVLRRENQPOIKVFNEVE
 PNPSTHTVYKGLEMFINFQPNIIALGGGSAMDAAKAIWMFFEHPETSFFGAKQKFLDIR
 35 KRTYKITKPKNAKFCIPTTSGTGSEVTPFAVITDSETHVKYPLADYALTPDIAIVDPQF
 VLSVPKDVAADTGMVLTTHAIESYVSMASDYTRGLSLQAIKLTFDYKSSVQENDKHSR
 EKMHNASTMAGMAFANAFLGISHSIAHKIGGEYGIPHGRTNAILLPHVIRYNAKDPQKHA
 LFPKYDFFRADTDYADIAKFLGLKGNTEELVDALANAVYDLGCSVGIDMNLKSQGVTEE
 40 LLHSTIDRMAELAFEDQCTTANPKEPLISELKGIIETAYDYER*

Sequence 755

Contig_0512_pos_8604_8128,

putative peptide of unknown function

45 gtgaataaggatgctgagaaccctaaacctaaagaaggatagggacttgattggaaaa
 gatattaaaacactaacgcattatggacaagctgatcggcttccatataaaaaat
 ggggttaaaaaattatgtcttttaacagaaagatgaatattatattgtaagtactaataaa
 ggaacaatcacatcagtttatgccacaggtaaaaggtgtgaaagtgaagccacttaaaata
 ggtgaaagttcatctcatatttttgaagatactagtattaatccagaaccaactgtcaaa
 50 acgaaaggtaaaacttataaatttgaaatgtctgatgaagacttaagacacagacgtta
 attaaatatggagatgtttatgctcaaatatattctgatcaacaaactaataaaattttg
 cgagtgaatttttagatgcaaatacattggcaacactacaacatataagttataa

Sequence 756

55 VNKDAENPKPKEGIGTWIGKDIKTLTHHYGQADRSYPYKNGLKNYVFKQKDEYYIVSTNK
 GTITSVYATGKGVKSPLKIGESSHIFEDTSINPEPTVKTKGKTYKFEMSDLDLKTQTL
 IKYGDVYAQIYSDQQTNKILAVRFLDANTLATLQPYKL*

Sequence 757

Contig_0512_pos_8113_7241,

is similar to (with p-value 4.0e-18)

>pir:pir|S58131|S58131 integral membrane protein LmrP - Lactococcus lactis >gp:gp|X89779|LLLMRP_1 L.lactis DNA for LmrP gene. NID: g1052753.

5 atggatgcgattacgcctgaagttgagcaatatatttataagataagttattggctgacg
aatattgctgtcgccttttggcgctcataggtggattgatgtatggggcacataaatct
atgttgtttttcattcgcttttggcatttacattatggtttttatagcacttatcgatgg
ttgcctaaagattttaaattattgttactcagtcgcacacacatcatgctaagagaaacaa
10 ttctccatgggtcaaatattaaaaagttataaaccagcatttaaagatacaacatatcta
cttctaattataggatttagtatttttaacaatgggtgagttatctgcatcatcgtaatt
tcagtgcgttttaaaacaagagtttgatccgatgatattgttttcgttacatatcaatggc
gttaaaatgtattcacttctattaatgacgaatacaatcattgttataatttttacctat
tttattttcaaaaattgttatgagaatgaatgttaaaacagcattattggttgaattatt
15 tttatgtcatttgatattcgaacctcacttatcttaatgattttacgttacttatcata
tttatgattatagcgacgataggtgaaatggatatattctccaattcttgaagaaaatcgt
tttaaaatggttccttctcataaaagaggacatatcagcagtgcatgcttttaggattt
aacctagctgaattacttgcaagatttggaattatattaggagtgttttaacttcaatg
gagatggggatctatatgtttgttttatattactaggtggcatgtcactttacattgca
20 gtgagtcgttttaataatacaaaattcacataa

Sequence 758

MDAITPEVEQYIYKISYWLNTNIAVAFGALIGGLMYGAHKSMLFFIAFVIYIMVFIALIVW
LPKDLNIVTQSHTHANEKQFSMGQILKSYKPAFKDITYLLLIIGFSILTMGELSASSYI
SVRLKQEFDPMILFSLHINGVKMYSLLMNTNIIIVIIIFYFISKIVMRMNVKTALLVGII
25 FYVIGYSNLTYLNDFTLLIIFMIIATIGEMVYSPILEENRFKMPVPSHKRGTYSAVHALGF
NLAELLARFGIILGVFLTSMEMGIYMFVLLLLGGMSLYIAVSRFNNTNSQ*

Sequence 759

Contig_0512_pos_0_6968,

30 is similar to (with p-value 0.0e+00)

>gp:gp|AF007865|AF007865_3 Bacillus licheniformis bacitracin
synthetase operon including bacitracin synthetase 1 (bacA),
2 (bacB) and 3 (bacC) genes, complete cds. NID: g2982193.

35 gtgacatactgggtaaagtttaagtcgcgacattgagttacgtagattaatgtatgcatta
ttagatgtcgttcaaaagtcacacctgtgttgcgtacacagtttgtgacagatgattttaat
caactcaagataaaatttaagagattttttccattttattgaaattaaagaagttaatgaa
atgtcgcaaaagcatagatttagaagcattctttacacgtaatttaaattcctaccatttc
aatcaattacctctgttttaattttaagatatatcaatttctagatggcgctacctaatt
40 ttagatttccacgctactatttttaattgaaagtcatttaactccatttttacaacaatta
aatattgcttataccactcttttaaaagtgaaatagatatctcggtttttataattgg
attaaagaaatgaatcaaaagatggatcaaaatcaagttgtgtgtccatcaaagcattc
aacgtattgaatgcagacggtgataattacgcttacataacctgtaaagaatacatctgaa
aagaaaaaaatgtgtctttgcatgcagaactaccatctttagacattgatgcgtggatt
45 gtaagttttacttagcgcatcattttataagtcagtcctctgatgtgacgttaggcac
catttttcgatagataataaaaaactgagaatatgatggtttttaaacacagacattgcc
ccacttaatttaagtattagtcaaagtgacgcgctaaaagatatggtggatgagtggtcc
gcgctacttgaagagcttcaaatgtgtggtgcgtctttgttgttcaacctaaagcagta
caaatagatgtagaacgatgattcatattgaaaaagtacaagaacaatttgagcttaatt
50 catatatgtcatatatacatcgtctatacaatgaagcatcatcattcgcggatttagag
ttttatcctcatgtgcaggggtggttttgatatagtttataatgacaacggtttatgatgaa
ttaactgtaaatacgttagtcaaattaattaatgggatttatatgcaaattacacaaaat
ccatcattatttaattaaagatataaaactcagtgatcgctcagatttagctaaatataat
gacatcaatcttcaaaacaatgacattaattatagtgaagtcacttataaaaccgtggtt
gaaagattcgaacgtcaagtgacccaacatcccgatagattgcgttgcaatatgaacaa
55 cgatcgatgacatatcatcaattaaatcaatgtgcgaatcttttagcatatagattgcgt
ttaaatcatcagattgaacctaatgatattggtggcatttaatagcagaacgcagcttagaa
atgattattggaatgttagggatcttgaaagctggtgcaggctacataccaattgatccg
gattatcctgaagaaagaatgaattatattattgaggacgcaaacctaaagcggttgta
acatatcgtacatcatttcaatcaggtttacctcaaatggatatagaattgatagttgat

tcaagagaacatgatattgataacccgagaggcattaattgttcagaagatatcgcttat
gtcatctatacatcaggaacgactggtaaacctaaagggacactgggtgccacatagagga
attgatcgcttagtacacaatccaaattatgtcgaattgaacgaaaatacaaccgtctta
ttatcaggaacagtagcttttgatgcagcaacctttgaaatatatgggtccattattgaat
5 ggtggacggttagtcattacatctaaagatagcttggttaaactcctcaattgttagatcaa
gctattactgaaaataaagtcaacacgatgtggtaacgtcatctttatttaatacaaat
gctagcgaacgatcgaagcactagaatctttaacttatttacttattgggtgggaagt
ttaaatgctaaatgggttcacttattaaattcgcgtagtgcacctcaataatcaat
ggttatggaccgacagagaatacaacatttactacaacttttgcgattccacaagagatg
10 ccttcacgtatacctattgggtttacctatttagtggaaacgacagtttatgtcatgcaaggt
aatcgattttgtggcgttaggtgttccaggtgaattgtgcattgggtgggtgcaggttagca
aaaggttatttaaatcaacctaaacttactgctgaacgttttattcagtcaccttttaac
aatgaaatgctttatcgaagcgggtgatttagttcgtcttcaagaagatggctatatgat
tatattagtcgtatcgataagcaagttaaaaacgcggttttagaatagaattatcagaa
15 attgaaaaagcattagaagctatacgtgatattaataaagctgtagtcatcggtcgagag
caagaccaagataaaacaaatagtggtcatattatgaagcatcgcaattaaaaatcaacaggt
caattaaaagatatatttaagtgaacattacctgaatatatgatacctgtgcattttatg
aaggtggatcgtatacctatcacgatgaatgggaaattagatgtgcgtgcattacctgaa
attaatctaaagaataatagaattatgtagaaccacgtaacgatattgaacgcacagtt
20 tgcgctattttcgaagagattttacatgttgatcaggtaggtgttaaagataatttcttt
gaactaggtggacactctcttagagcaacattagttgtaaacggtattgaagaaagggtta
aaaaaacgtcttaagtaggtgatttaataaagcgcctactgtagagcaacttgacaa
caaatggaagaactgcaaaatgatgtctatgaagtgattcccaaagcaaatgaatcgat
caatatgatttaagtgcgtctcaaaaaagtatgtatctttatggaaggtcaatcctaaa
25 gacacagtgataacattccattcttatggagattatcttctgaacttaattgttatgcaa
ttgcaacgtgatttactaagttgattgaacgtcatgaaatattacgaacacaatatgta
attgatgacaatgaagttaaacaacgtattgcgacacatgtttcgctgattttgaagag
gtaacgacatctctaaacgaacgagcaagatattattcaatcatttatggaaccgtttgat
ttagaacaaccaagtcagatgcgagttaaatatatcatggaccacaacaagattattta
30 ttatggatactcatcatagttatattatgatgggtatgagtaaacacgattttactatctgat
ttgaacgctttataccaagataaaatcattacctgaacttaagcttcagtataaagattat
agtgagtggtgggtgcacagagacttatctaaacaacgtcacttttggttacagcaattt
gaaaatcaggttccaatatataaatatgcctacggtatttcttagaccaagttataaaaca
accaacggtaatatgttgacgtttcattacaatcgtcaaatcaaacagcaattgaaatct
35 tatgtagaacaacatcaagtgacagactttatgttctttgctagtgcattcatggtatta
ttgcacaaatatcacgctcaggacgatatcgctattggtagtgtaatcagtcgcggtact
catcgcgatactgaaaatatgttaggtatgtttgctaatacacttgtatatcggtgga
ccacatgatcaaaagacatgggatcaattgatggctgagatgaaagagatgtgtctaggg
gcatatgaacatcaagaatatcctttgaaagcttagtcaatgatcttggtagaaga
40 gatgcttcacataatccgttatttgatgtgctcgtacttcaaaataatgaaacaaat
catgcgaattttggacatagtcaattgacacatattccacctcagtcacaacagctaaa
tttgatttgcatttattattgaagaagatcaagatgactatgtcgtcaatattgaatat
aatacagatttatataaacaagagaccattcatcatattgctgaacaacttcaaatgatt
atataacatgtaatatctaccgaaaacctaataaattcaagatatattgatgaaatgatgac
45 ttatttaatttggttggaacgatgtgaattgttcttttagacttgccaaaaataaag
tcaatacagcaacttttacatgatgtcatgaaagcgaaagcagatgatgtagcacttaaa
atgaatggacaatcgatgacgtatcaagaacttgatgattattctaatagtatgggtcaa
acattgatatacaaaatggcattcaaaaagggaacgtgtagcccttttaactgaacgaagt
tttgaaatgggttgctagtatgattgctgtattaaaagttggaggttcttatgtacctatt
50 gacgtcacttatcccgataaacgcattgaatttattattgaagacgctgaagtcgcagca
gtgctcacatatggaaaagcaatatcctcacatataccagtaattaaaattgaagatatt
gataacactgaaaaataataaagggttaaatatagaatatgcagggaatttggaagatgat
atgtatcatatttatacatctggaacaacaggaaagcctaaagcagtatcagtgaaacaa
cgtaatatattaaatttagtatgtgcttggaacaaaagactcaatttatccgatgatgaa
55 gtctatctgcagtagcgttaattatgtgttcgatgcttcggcaactgatttctactgtagt
ttattaaatggatatccgcttgctcattgcaacatcagttgagcggaccaatacagattta
ttagaaaagtttaattcacaagaaaatatcaccatcgcatctattccactacaggtatat
aatgtgatgcacatttctatatccctaaagtgattacaggaggtgcgccaagtactcca
gcatttgttcaacatatcttaagcattgtgatgtacgttaatgcctatgggccttct

5 gaaaatacagttataacatcttgttggatatacgaaaaaggtgacgccataccatcgact
 attccgattgggaaaccggttagctaattgtgatatttttatatgtcaggcggttaaacta
 tgtggcggttggtattccagggtgaattatgtattgcaggagaaagtttaacttcaggatat
 ttaaacagacccgaactttctgctgaaaaatttataaataatccttttgggccaggacaa
 10 ctttatcgaaagtgggtgatttagcacgattgatgccagatgggcaaattgaatttcttgggt
 agaattagacaagcaagttaaagtacatggctatcgcatgaaactaggtgaaattgaaat
 atcattaattcagtagatactgttacagatagcggttgttattttagctaaacagggtgag
 cgtgaagtgtctgcatgcttattatgttggaaagtcaagaagatgaaagtcataattcacaa
 catttaaatcaatatttgctaaatacatgattcctaagacattaacagctatttagcgaa
 15 attccattaacaggaaatgataaggtggatgagtcagattacctgtacctaattgtacac
 aaaaaataatttgttgcaccacgtaataatatcgaacgagaaatagcacaatcgtagc
 ggagtgttggacgtatcgctctatgagtatagatgatgacttctttgaaatgggtggtaca
 tctactagatgctatgggtggttagtatacaaaactaaaatcaaatggcatacacattacaatg
 caagatgtatatcaatttaaaactgttcggttatatagctaatacacacagaaaaacgcaa
 20 gcactaccagaagtagtattaccagatcatctaccacaattacaatctttggttgaaaga
 cgatccaactaaaatcacacacctaacgcaatcatcttaggtcatgtattgctaact
 ggtgcaacaggggttcttaggcgcatatttaattgatgaaatgcaagatgatgctgatcaa
 attacatgtattgtcagaggtcatgatataatcaagctaaaactaacttgaaaaataat
 ttaaatgtttatttgcacggctcatgtggataaattaatgaagcacattgatattatt
 25 tttagcggatttatcagaacttgacctcttattatcgattcagccattgatacaattatt
 catgctggagctcgtacagatcactttggcgatgatgaaacatttttcgatgtcaatgta
 agaagtacacaagcatttaattgatttagctaaggataaaaaagcgaattaatctatata
 tcaacgataagtggtggtacggtatttgaagtacatcaagacgatattacatttctgaa
 aaagatttatataaaaggccagttatttacatcaccatacactaaaagtaagtttatagc
 30 gagattaaagtgttagaagcgggttaattgaaggtttagcagctcagattataagattagga
 aatctgacaagtgcttctactggaccattaaatatgaaaaatttaacaactaatcgtttt
 agtattgtcatgcatgatttattaaaaatgccggtttataggagaaagtatatcgaaagct
 aaagttgaattttcatttatcgatgtcacagcgccalattattaaattggcaagatcc
 aatgcaatacctattatttatcatgtatacgccatgttcgataactatgaaacaagta
 35 attgacaatgccaaagggtcagaaatgactgtagtaagtgatagtgagttgaacagaaa
 ttacatgaattaggtatgcatgaattgatgtgtcttaatagtaattggagataatcaaatt
 tcagggtgt

Sequence 760
 35 VTYVWKLSDIELRRRLMYALLDVVQSQPVLRTQFVTDDFNQLKINLRDFFPFIEIKEVNE
 MSQSIDLEAFFTRNLNSYHFNQLPLFNFKIYQFLDGAYLLDFHATIFNESQLTPFLQQL
 NIAYTHSLKSEYSISDFYNWIKEMNQMDQNVVCPKSHFENVLNADGDNAYIIPVKNTSE
 KKKMCSLHAELPSLIDAWIVSIYLAHHFISQSSDVTGLIHFSIDNKNTENMMVLNTDIA
 40 PLNLSISQSDAVKDMVDECSALLEELQMGASFVVQPKAVQIDVETMIHIEKVQEQLFELN
 HICHHIRLYNEASSFADLEFYPHVQGGFDIVYNDNVYDELTNTLVKLINGIYMQITQN
 PSLLIKDIKLSDRDLAKYNDINLQNDINSEVYKTVVERFERQVHQHPDSIALQYEQ
 RSMYHQLNQCANLLAYRLRLNHQIEPNDMVALIAERSLEMIIGMLGILKAGAGYIPIDP
 DYPEERMNYIIEDAKPKAVVYRTSFQSGLPQMDIELIVDSREHDIDNPRGINCSEDIAY
 45 VIYTSGTTGKPKGTLVPHRGIDRLVHNPNYVELNENTTVLLSGTVAFDAATFEIYGPLLN
 GGRLVITSKDTLLNPQLLDQAITENKVNTMWLTSSLFNQIASERIEALESITYLLIGGEV
 LNAKWVHLLNSRECHPQIINGYGPTEMTFTTTFAIPOEMPSRIPIGLPISGTTVYVMQG
 NRICGVGVPGLCIGGAGLAKGYLNQPKLTAERFIQSPFNEMLYRSGDLVRLQEDGYID
 YISRIDKQVKIRGFRIELSEIEKALEAIRDINKAVVIVREQDQDKQIVAYYEASQLKSTG
 50 QLKDILSETLPEYMPIPVHFMKVDRIPITMNGKLDVRLPEINLKNRNYVEPRNDIERTV
 CRIFEEILHVDQGVKDNFFELGGHSLRATLVVNRIEERLKKRLKVGDLMKSPTEQLGQ
 QIEELQNDVYEVIPKANESYQYDLSASQKSMYLLWKVNPKDTVYNIPLWLRLSSELNVMQ
 LQRALSKLIERHEILRTQYVIDDNEVKQRIATHVSPDFEEVTTSLTNEQDIIQSFMPEFD
 LEQPSQMRVKYIHGFPQDYLFMDTHHSINDGMSNTILLSOLNLYQDKSLPELKLQYKDY
 SEWMVHRDLQKQHFQFENQVPIINMPTDYPRPSIKTTNGNMLTFHYNRQIKQQLKS
 55 YVEQHQVTDFFMFFASAIMVLLHKYTRQDDIAIGSVISARTHRDTENMLGMFANTLVYRGR
 PHDQKTWDQLMAEMKEMCLGAYEHQYPFESLVNDLVDERDASHNPLFDVMLVLQNNETN
 HANFGHSQLTHIPPQSTTAKFDLSFIEEDQDDYVYVNIYNTDLYKQETIHIIAEQLQMI
 IKHVISTENLKIQDIDENDLLIWLKHNDCSLDLPKNKSIQQLLHDMKAKADDVALK
 MNGQSMYQELDDYSNSMAQTLIQNGIQKGERVALLTERSFEMVASMIAVLKVGGSYVPI

DVTPDKRIEFIIEDAEVAAVLTYGKAISSHIPVIKIEDIDNTENNKRNLNIEYAGNLEDD
 MYHIYTSGTTGKPKAVSVKQRNINLVCAWTKRLNLSDDDEVYLQYANYVFDASATDFYCS
 LLNGYPLVIATSVERTNTDLEKLISQENITIASIPLQVYNVMHFFYIPKVITGGAPSTP
 5 AFVQHISKHCDMYVNAYGPSENTVITSCWIYEKGDAPSTIPIGKPLANVDIFIMSGGKL
 CGVGIPGELCIAGESLTSGYLNRPESAEKFINNPFPGQLYRSGDLARLMPDGQIEFLG
 RIDKQKVHGYRIELGEIENIINSVDTVTDSVVILAKQGEREVLHAYYVGSQEDESISQ
 HLNQYLPKYMIPKTLTAISEIPLTGNDKVDSESLPVPNVHKNKFVAPRNNIEREIAQIVS
 GVLDVSSMSIDDDFFEMGGTSLDAMVVVSKLSNGIHITMQDVYQFKTVRYIANHTEKRQ
 10 ALPEVVLDPHLPQLQSLVERRYQLKSQHLTQSSSLGHVLLTGATGFLGAYLIDEMQDDADQ
 ITCIVRGHDINQAKTNLENNLNCYFDTAHVDKLMKHIDIILADLSELDHLIIDSAIDTII
 HAGARTDHFDDTEFFDVNVRSTQALIDLAKDKKAKLIYISTISVGTVEVHQDDITFSE
 KDLYKGQLFTSPYTKSKFYSEIKVLEAVNEGLAAQIIRLGNLTSASTGPLNMKNLTNNRF
 SIVMHDLLKMPFIFGESISKAKVEFSFIDVTARHIIKLARSNAIPIIYHVYAPCSITMKQV
 IDNAGGSEMTVVSDSEFEQKLHELGHELGNSNGDNQISGV

15 Sequence 761
 Contig_0513_pos_522_938,
 putative peptide of unknown function
 atgttcccacccgaacacctagtagagatgccactaacccacctcaacaactcttataca
 20 ttactaggaataaccgtcaaatattacccatctcacatacaatttttctgagcatgcatta
 ccttgataagttttatcgtagactatagttttctatcgctattgcaataatctatatt
 tatatcgcaagaaatatacaaaaatcacactagggtatggtgctttatttggtatagtt
 atttgattgtttttcatttaattcctaattgcaattatgcatgctgacgaatgctttt
 gatcaaccattttcagaacacctatcagaattttttggacacattgtttggatgattacg
 25 tctacgtacaacacgcattttcattttttgtcggtgttttttcttattttcttgtga

Sequence 762
 MFPPRTPSRDATNPPQQLQLLGIPSNITHLTYNFSEHALPWISFIVHYSFSIAIAIIYI
 30 YIAKKYTKITLGYGALFGIVIWIVFHLILMPIMHVVPNAFDQPFSEHLEFFGHIVWMIT
 STYNTHFHLFLFLFL*

Sequence 763
 Contig_0513_pos_10290_9811,
 putative peptide of unknown function
 35 atgagtagtgtagtaacacttaaggcttacccttagcttataataaagacatgcaagaa
 gataaagaaggtttatttgatgctgtacacacacttaaggctctcttcgaatcttcgaa
 ggtatggttgcatctatgaaagttaattcaaaccgtttaagtcaaacagtaaaaaatgat
 tttcaaatgcaacagaattagcagactattagtcagtaaaagtgtaccttttagaacc
 gctcatgaaatcggttgtaaaatcgattaaattgtattcataaaggatataacctatta
 40 gacgtacctttaagcgaatatcaagaacatcatgagaatttgaggaagatatatatgat
 tatttaacacctgaaaattgtctcaagcgtcgcaaaagctatggttcaactggtcaagaa
 tcagtaaaacatcaactaaaagtcgcaaaagcattattaaaagacaacgaatcaaaatag

45 Sequence 764
 MSMLVTLKGLPLAYNKDMQEDKEGLFDAVHTLKGLSLRIFEGMVASMKVNSNRLSQTVKND
 FSNATELADYLVSKSVPFRTAHEIVGKIVLNCIHKGIYLLDVPLSEYQEHENIEEDIYD
 YLTPENCLKRRQSYGSTGQESVKHQLKVAKALLKDNEK*

50 Sequence 765
 Contig_0513_pos_9403_9059,
 putative peptide of unknown function
 gtgacaaaccggaggaaggtggggatgacgtcaaatcatcatgcccttatgatttgggc
 tacacacgtgctacaatggacaatacaaaagggcagcgaaactgagaggtcaagcaaattcc
 55 cataaagttgttctcagttcggttagtctgcaactcgactatatgaagctggaatcg
 ctagtaaatcgtagatcagcatgctacggtgaatacgttcccggtcttgtacacaccgcc
 cgtcacaccacgagagtttgtaacaccggaagccggtggagtaaccatttgagctagcc
 gtcgaaggtgggacaaatgattggggtgaagtcgtaacaaggtag

Sequence 766

VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSSDCSLQLDYMKLES
LVIVDQHATVNTFPGLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVTR*

5 Sequence 767

Contig_0513_pos_5415_3979,

is similar to (with p-value 0.0e+00)

>gp:gp|AF054624|AF054624_1 Lactobacillus sakei transcription
-repair coupling factor (mfd) gene, partial cds; L-lactate d
ehydrogenase (ldhL) gene, complete cds; and unknown genes. N
ID: g3511014.

atgcaagattttccggtcgaaattcaattggttaagtcgattccgcacagctaaagaaata
agggaactaaagaagggtctcaaatcaggatatgttgacattgtcgttaggtacacataaa
ttattaggtaaagatattcaatataaagatttgggattgcttattgttgatgaagaacaa
15 cgttttggagtgcgacataaagaacgcattaaaaactttgaaaaaaaaacgttgatgtactg
acgcttactgcaacaccaataaccaagaacattgcatatgagtatgttaggtgtacgtgac
ttatcagtgattgaaacaccacctgaaatcggttttcctgtacaaacttatgtcttagaa
cagaatacgaactttattaaagagcgattagagcgatgaattatctcgcgatggacaagta
ttttatttgtatacaaaagtgcagtcattttatgaaaaagagaacaacttcaaagggtta
20 atgcctgacgctaacattgctgtagcacatggccaaatgactgaacgtgatttagaggaa
acaatgttaagctttattaatcacgagtagcatatttttagtaacgactacaattattgaa
acagggtgtagatgtaccaaattgctaatactttaatcatagaagaggctgatcggtttggt
ttaagccagctataccaattaagaggacgtgtaggacgttcaagtagaattggttacgct
tatttcttacatccagctaacaagtggttaaatgagactgctgaagagcgattgcaagct
25 attaaaggagtttacggaactaggttcagggttttaaaatcgctatgagagatttaaatatt
cgtggtgcaggcaatttactcggttaagcaacaacatggctttattgattcggttggtttc
gatttatactctcaaatgttagaagaagcagtaaacgaaaaacgtggcattaaagaagaa
tcgccgatgcaccagatattgaagtagaattgcacttagatgcttattaccagctgaa
tatatacaaaagtgaacaggctaaaatttgagatttataaaaaacttcgaaaagtagaaact
30 gaagaacaacttttcgatgtcaaagatgaattaatagatcggttttaattgattatccaatt
gaagtgcgaacgattattagatattgttgaaatcaaaagtcacgctctacatgcgggtgct
gaattgataaaagacaaaggcaaatctatacaaatcattttatcacctaaagcgactgaa
gatattaatggagaagaattgtttaaacagacgcaacctcttggttagagcaatgaaagtt
ggcgtgcaaaataatgcaatgaatgaacgctaacaaaatcaaaacaatggtagatagt
35 ttgaaattcttagttagatgtattgaagaaagtatggcgattaaagatgaagactaa

Sequence 768

MQDFPVEIQLVSRFR TAKEIRETKEGLKSGYVDIVVGTHKLLGKDIQYKDLGLLIVDEEQ
RFGVRHKERIKTLKKNVDVLTATPIPRTLHMSMLGVRDL SVIETPPENRFPVQTYVLE
40 QNTNFIKEALERELSRDQGVFYLYNKVQSIYEKREQLQRLMPDANI AVAHGQMTERDLEE
TMLSFINHEYDILVT TTIETGVDVPNANTLIIEEADREGLS QLYQLRGRVGRSSRIGYA
YFLHPANKVLNETAERLQAIKEFTELGS GFKIAMRDLNIRGAGNLLGKQQHG FIDSVGF
DLYSQMLEEAVNEKRGIKEESPDAPDIEVELHLDAYLPAEYIQSEQAKIEIYKKLRKVET
EEQLFDVKDELIDRFNDYPIEVERLLDIVEIKVHALHAGVELIKDKGKSIQIILSPKATE
45 DINGEELFKQTQPLGRAMKVGVQNNAMNVTLT KSKQWLD SLKFLVRCIEESMAIKDED*

Sequence 769

Contig_0513_pos_3926_2448,

is similar to (with p-value 6.0e-48)

>sp:sp|P37555|YABM_BACSU HYPOTHETICAL 57.4 KD PROTEIN IN MFD
-DIVIC INTERGENIC REGION. >gp:gp|D26185|BAC180K_120 B. subti
lis DNA, 180 kilobase region of replication origin. NID: g46
7326. >gp:gp|Z99104|BSUB0001_57 Bacillus subtilis complete g
enome (section 1 of 21): from 1 to 213080. NID: g2632267.
55 gtgaagataactaagtgccatttatcgattccgtatcaaaatgttttaggtgatgacggt
ttatatgcttatacaaaatatatcctgtcgtagcactaggggttattttatctatgaat
gctattccaagtgtgtgactcaagtgataggtgttaatcgatccgatgaagtctataca
agggttatgtttcgattacaatgcataggttttatcgctcttattttgctttttatgttt
gcgaatatgattaccgatggatggcgattctaattagcacccatgttaaagatggcc

agtttttagttttattttaaataggtgtcttaggagtgttaagaggattttatcaatcaaaa
 caagtaatgaccataccagcaatttcccagggttatagaacaggtaattagagttagttta
 atcattgttgcaattattatgttttcaatgaaacactggctctatttatcaagcaggagca
 ttagctatattggcatcttcgattgggttttttaggttcaatgttatatttattacttaaa
 5 aaaccacttaaaacttaagttatgctatcgctttaataataacttccattcaatggaagcag
 ttgtttatttccatatccatatttgcattgagtcacattatcggtattttatggcaaggt
 gtggatagttttacaataatcggtttattacaacatagcgggtattgcttttaagaagca
 attattcaaaaaggcatttatgatcggtggtgcttcattatacaaatgggtttgattgta
 actacgacttttagtttcttctatcccattacttactcaagcaattcggtgaacataat
 10 caaattcatatgaatcggttatgcaaatgcatcaattaaaatcacggtagtaataagtaca
 gcagctagtataggattaattaatctgcttccacttatgaatgttgattctttaaaagt
 aatcatttaactctaactttgagtggttatatgtttacagtgatattgtttcggttaata
 atgatgaatatctcattattacaaggtcaaaccaggtattcgctccattattatgggtgtg
 ataataaggaatactgtccaaaatttttaaatgttatattaatacctttttgggggtatc
 15 gtgggtgcaagtgtagtacagtccttatcactactactttttgtcataatattgcaagtt
 gcagtcctaaagtactaccgttttaaatcgatatatctttatttatcggttaaacttatttta
 ggtatgataattatgagtatagttgttcaaactgtcatgcttgccctaccttcaaaaagt
 aggatgttaggattactagaacttatagtttagctcaattataggcatagtgattataatg
 ttgtatattattatatttaaatgtattaggatacaaaagaataaagcacttaccttttggg
 20 gacaaatttatcaaatgaagagaggaagacgggtcatga

Sequence 770

VKILSAIYRIPYQNVLGDDGLYAYQQIYPVVALGVILSMNAIPSAVTQVIGVNRSEVYT
 RVMFRLQICIGFIVFILLFMFANMITRWMGDSNLAPMLKMASFSFILIGVLGRFYQSK
 25 QVMTIPAISQVIEQVIRVSLIIVAIIMFSMKHWSIYQAGALAILASSIGFLGSMYLLLLK
 KPLKLLKLCYRFNNTSIQWKQLFISISIFALSQILIVLWQVDSFTIIRLLQHSIGIAFKEA
 IIQKGIYDRGASFQMGILIVTTTFSEFLIPLLTQAIREHNQIHMNRYANASIKITVVIS
 AASIGLINLLPLMNVVFFKSNHLTLTSLVYMFTVICVSLIMMNISLLQVQTSIRPIIMGV
 IIGILSKIILNVILIPFWGIVGASVSTVLSLLLVIILQVAVLKYYRFNRISLFIVKLIL
 30 GMIIMSIVVQTVMLALPSKSRMLGLELIVSSIIGIVIIMLYIIIFNVLYGKEIKHLPFG
 DKLYQMKRGRS*

Sequence 771

Contig_0513_pos_2112_1261,
 35 is similar to (with p-value 3.0e-60)
 >sp:sp|P37556|YABN_BACSU HYPOTHETICAL 56.1 KD PROTEIN IN MFD
 -DIVIC INTERGENIC REGION. >gp:gp|D26185|BAC180K_121 B. subti
 lis DNA, 180 kilobase region of replication origin. NID: g46
 7326. >gp:gp|Z99104|BSUB0001_58 Bacillus subtilis complete g
 40 enome (section 1 of 21): from 1 to 213080. NID: g2632267.
 gtgaaggtacttgaggaaaaagttttattgatgacatttttgaagcgggtgatgtagac
 cccaatgatgggttttacactgcttgatgggtacgtcattaaaagaatcgccctaaatgtt
 cgtacaaatacagtaattactcaagtttatagcgtaatgatagctgccgattttaaactt
 actttaatggaaagatatcctgatgattttaatgtgaaaataattactggttctcatagt
 45 gatggagctcacgtaattgaatgccactttatgaaattgatcgctacgacgattatatt
 aataatcttacaagttttatttattccaaaaatcaatgaggatacattactttatcaagat
 tttgattacgcagttcaaactattgatttactcggtgataatgaaaaagggtgtccgtgg
 gataaagtacaaactcacgactcattaaaacgggtatcttttagaagaaaacggttgaatta
 tttgaagccattgataatgaagatgattggcatatgatagaggaattaggagatatactt
 50 ttacaagtattattacattctagtataggtaaaaaagaaggatatattgatatcaaagaa
 attatagaaagctcaacaccaagatgattcatagacatccacatatcttggtaatgcg
 catgtaacttcgaagaggattttaaagacatttgggtcacgtgctaaagaaaaagaaggt
 aaagtgcctcggtttaaatttgagaaagtatttgcagaccacttcttgaattgtatgat
 aaaacaaaaaataggcaatttgacgaagatgatctcaacaatttttacaacaaggagag
 55 aaaaattcatga

Sequence 772

VKVLGGKSFIDDIFFAVDVPNDGFTLLDGTSLKESALNVRTNTVITQVYSVMIAADLKL
 TLMERYPDDFNVKIIITGSHSDGAHVIECPLYEIDRYDDYFNNLTSLFIPKINEDTLLYQD

FDYAVQITIDLLVDNEKGCPWDKVQTHDSLKRYLLEETFELFEAIDNEDDWHMIEELGDIL
LQVLLHSSIGKKEGYIDIKEIIESLNTKMIHRPHIFGNAHVTSQEDLKDWSRAKEKEG
KVPRVKFEKVFADHFLKLYDKTKNRQFDEDDLKQFLQGEKNS*

5 Sequence 773

Contig_0515_pos_5400_6044,

putative peptide of unknown function

atgtctaaaatgcaaaactgccccaaacaatgattatcatgcaagtatctatctttatcatc
gctgcattgttaatcattttcatgcaagcaaccattaaaaatcctactcaattagaacaa
10 ggacataaggaacctaaaagatatatcttcgcctgggttctattaggattttgtattgtg
atgattttaccaagtcacatcagtaataatctttatggccatcaatggcagtcgcgaaaga
agtccaaacactgaaagggttaatggctattgctaagcaaatgccgatatctatcggttta
atatccatagtaggtcccatttttagaggaatacgtttttcgtaaagttatcttcggagaa
ttgtataactttataaagggtacagtggtgtaagctttatcattgcttctatagtaagt
15 tctctaataatcgcttttagcacataatgacttcaaattttataccggtatattttggaatg
ggagtcattttctcacttgcttatgtttacacaaaacgaattgctgtacctataggcata
cacatgcttatgaatgggttcggtgtattaactcaagttgtgggtggagattctattaaa
aaattgcaagaacaagcaacatttatattccatcttatattttaa

20 Sequence 774

MSKMQTAQTMIIIMQVSIFIIAALLIIFMQATIKNPTQLEQGHKEPKRYIFAWVLLGFCIV
MIYQVVISIILFAINGSPQSPNTERLMAIAKQMPIFIVLISIVGPILEEYVFRKVIFGE
LYNFIKGSRVVSFIASIVSSLIFALAHNDFKFI PVYFGMGVIFSLAYVYTKRIAPIGI
HMLMNGSVVLTQVVGDSIKKLQEQTFFIHLIF*

25

Sequence 775

Contig_0515_pos_7937_8731,

is similar to (with p-value 0.0e+00)

>gp:gp|AF012132|AF012132_6 Staphylococcus epidermidis agr sy
stem including response regulator (agrA), histidine kinase (
agrC), AgrD (agrD), AgrB (agrB) and delta toxin (hld) genes,
complete cds. NID: g2981293.

gtgaatatattgaaaatccaaataacttcaattcaatgtagaacgtggaaatgttgataaa
aatatgcaaaaatatcaaaaactaagtttaataacttagataaaagataaccagtgctcgtc
35 gtgcttcagaaatgtggaataacgggttatgcattagaagaattagaacaaaaagctgat
aaaaatcttaaaagacagctctctctttataaaaagacttagcacatacatttaatgtagat
atcattgcaggttcagtggtcaaatataagagaaaacatatataataactgcttttgca
attaataaaaaacaaagaattgattaatgaatatgacaaagtacatctcgtgccaatgtta
cgtgagccagactttttatgtggtggaaatgtagtccctgaacctttttatttatctgat
40 caaacacttttgacgcaaatcatttggttatgacttgcgatttcagagatattgcgctat
ccagctagaaaaagggtgctaaaaattgctttttatgtagcgagtgccctagctcaagacta
gatcattgggttatcattactaaaagcgagagcaatcgaaaatgatatttttattgtagct
tgtaatatgttggtgatgatggtcacaccaattatgctggaaattcaattgtcattaat
cctaattggtgaaatttttagaccatttagatgataaagaagggtgtactaacaacacatatc
45 gatgtagacttagtagatcaacaaagagaatatattccagttttcagaaatctaaaacca
catctttataaaatag

Sequence 776

50 VNILKIQILQFNVERGNVDKNMQNIKTKFNQYLDKDTSVVVLPEMWNNGYALEELEQKAD
KNLKDSSLFIKDLAHTFNVDIIAGSVSNIRENHIYNTAFAINKNKELINEYDKVHLVPL
REPDLFCGGNVVPEPFYLSQTLTQIICYDLRFPEILRYPARKGAKIAFYVAQWPSSRL
DHWLSLLKARAIENDIFIVACNSCGDDGHTNYAGNSIVINPGEILDHLDDEKGVLTTHI
DVDLVDQQREYIPVFRNLKPHLYK*

55 Sequence 777

Contig_0515_pos_9479_10069,

is similar to (with p-value 5.0e-45)

>gp:gp|Z49220|SEHL DGN_2 Staphylococcus epidermidis hld and a
gr[A,B,C,D] genes. NID: g3320006.

atgacacttgaggagagtagaaaaacaagtgaaaatcatcgataaaaaaattgagcaattt
 gctcaatattttacaacgtaaaaataacttagatcacattgagtttctaaaagttcgttta
 gggatgcaagtagttgctggttaattgaaaaaacagtggttctatatggactatcttat
 ttttttgatttgctcatttttacatttttaactcatattagttactttctattaagaata
 5 tttgctcatgggtgcacatgctaaaaacaactcttcaatgctatatacaaaatattctttat
 tttttatttttaccttggttagtactacaccttcctcttagtacaaaatattttatttt
 ttagccatgattagttttttattagtaatatcttttgacccggctgcaacaaagaaacaa
 cctatacctaaacgtttacttaagaagaaaaaagtactctccatattaagttttattgta
 10 atcataacaatcgctttaacactagaagaagtattcaaaaaaatgttatctcgggtgtt
 gtaatagagtctattacacttttaccaatattttttcctaaggaggattaa

Sequence 778

MTLEESRKQVKIIDKKIEQFAQYLQRKNNLDHIEFLKVRMGQVAGNIEKTVVLYGLSY
 FFDLLIFTFLTHISYFLLRIFAHGAHAKTTLQCHIQNILYFLPLVHLPLSTNIFYF
 15 LAMISFLLVISFAPAATKKQPIPKRLKKKKVLSILSFIVIITIALTLEEVFKKNVISGV
 VIESITLLPIFFPKED*

Sequence 779

Contig_0515_pos_10231_11529,
 20 is similar to (with p-value 0.0e+00)
 >gp:gp|AF012132|AF012132_2 Staphylococcus epidermidis agr sy
 stem including response regulator (agrA), histidine kinase (
 agrC), AgrD (agrD), AgrB (agrB) and delta toxin (hld) genes,
 complete cds. NID: g2981293.
 25 gtgtattcaatgggtaaaacttgactttttaccattttgcagctatacaagtgtttcttttg
 gtttgggttacaaaaactattgctaattattaaatttgtaagaaaggattatattttcatt
 actggaattataatcctttctgcaatattatataatgtttatgcaagccaagcacttgta
 cttgtagtaataatgattataattttcttctattcaaaagtaagatggatttctattgtt
 atagtgttaatgagcactttgtgtcatatttaacaaattttattacagtagctatcagt
 30 ttataactgaaaatataataacataatattttattttataatattcttcaatttttcaatt
 tttatcattttatctttaattttggcacattttattaaacactttattaattaggtttagg
 tattcttattttatatttaagcaaaaggattacattattattttctttgtattagctatt
 gctttgatataacttctatataatttcacaaactaatttacaagaaagtaatagcttgaa
 ttttatgctattattttgtttctattaccgtacttttgagtttggttatattattgtta
 35 tcggctttcgcactacgtgaaatgaaatataaacgtaagctacaagaaatcgaagcatat
 tatgagtacacgttacgtatagaaagcattaacaatgaaatgcgtaagttccggcatgat
 tatgtgaatatcctcaccactctttcagattacattagagaagatgatatgcctggatta
 cgtaaatatttttaatgaaaatatcgttccaatgaaagataaattaaaaactcgctctatt
 aaaatgaatgggtattgaaaagttgaaagtgagagaaattaaagggtgattactactaaa
 40 attattcaagctcaagaaaaacgtatttccaattagattgaggttcctgatgaaattgat
 cgtatctctatgaatactgttgagcttagtcgtattatcggtattatagttgataatgct
 attgaagcttcagaaaatcttgaggaaccactcatcaatatcgcatcgcataatgag
 gaatctgtcactttttatcggttatgaataaatgtagtgtgatatccctaaaattcatgag
 ttgtttgaacaaggtttttctactaaaggtgataatcgcggtttagggtttatcaacttta
 45 aaagaactgacagactcaaacgagaatgttttattagatactgtcatcgaaaatggttac
 tttgtacaaaaagtagaaaataataaaggaaatcataa

Sequence 780

VYSMGKLDLFPFAAIQVFLVWVTKTIANIKFVRKDYIFITGIIILSAILYNVYASQALV
 50 LVVIMIIFFYSKVRWYSIVIVLMSTLLSYLTNFITVAISLYTENIHNIIYFYNIFHFSI
 FIILSLILAHFLFKHLLIRFRYSYLYLSKRYIIISFVLAIAFIYFYIISQTNLQESNSLN
 FYAIIFVSITVLLSLVILLLSAFALREMKYKRKLQEIEAYYEYTLRIESINNMERKFRHD
 YVNILTTLSDYIREDDMPGLRKYFNENIVPMKDKLKTRSIKMNGIEKLKVREIKGLITTK
 IIQAQEKRIPISIEVPDEIDRISMNTVELSRIIGIIVDNAIEASENLEEPLINIAFIDNE
 55 ESVTFIVMNKCSDDIPKIHLEFQGFSTKGDNRGLGLSTLKELTDSNENVLLDTVIENGY
 FVQKVEINNKE*

Sequence 781

Contig_0515_pos_11636_12262,

is similar to (with p-value 0.0e+00)
 >gp:gp|AF012132|AF012132_1 Staphylococcus epidermidis agr system including response regulator (agrA), histidine kinase (agrC), AgrD (agrD), AgrB (agrB) and delta toxin (hld) genes, complete cds. NID: g2981293.
 5 atggaggttagctttagcaacaaatgatccttatgaggtcttagagcaatcaaaagaactt
 aatgacattggtgttacttccttgatattcaattagaagctgatatgaacggtattaaa
 ttagccagtgaattcgtaaacatgatcctgttggttaattatattttgtaaccagtcac
 agtgagctgacttatttgacgtttgtttataaagtggtgctatggattttatttttaa
 10 gatgatccatctgaattaaaaatgagaatcatagattgtctcgaaacagcacatacacga
 ctcaaattattatcaaaagaaagtaattgtagatacgattgagttaaagcggggaagta
 tcagtatacgttcaatatgatgatattatgtttttgaatcatctacgaaatctcatagg
 ctcatgtcacatcttgataatcgacaaattgaattttatggaaatttaaaggaattagca
 cagcttgatgaacgtttcttttagatgtcataacagttttgtgataaacaggcataatatt
 15 gaatctattgactcaaaagaacgtattgtttactttaagaatggcgaaaattgtttcgct
 tcagtacgtaattgttaaaaaaatataa

Sequence 782
 MELALATNDPYEVLEQSKELNDIGCYFLDIQLEADMNGIKLASEIRKHDPVGNIIFVTSH
 20 SELTYLTFFVYKVAAMDFIFKDDPSELKMRIIDCLETAHTRLKLLSKESNVDITELKRGSN
 SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELAQLDERFFRCHNSFVINRHNI
 ESIDSKERIVYFKNGENCASFVRNVKKI*

Sequence 783
 25 Contig_0515_pos_13997_13296,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q05936|SCRB_STAXY SUCROSE-6-PHOSPHATE HYDROLASE (EC 3
 .2.1.26) (SUCRASE) (INVERTASE). >pir:pir|A47059|A47059 sucra
 se ScrB - Staphylococcus xylosus >gp:gp|X67744|SXSCRBA_2 S.x
 30 ylosus scrB and scrR genes. NID: g949973.
 atgataggagatttaaactttaataatctatttttcgaccatgaaagttttcaagaattg
 gataatggttttgatcttacgcgccacaaacggttggtgatgcagacgggcaacgcatt
 ttaattggatggatgggactaccagatacagagtacctacagataaagaggggtgggca
 cattgccttactattcctcgagtacttaccattgaaaatggaaaacttaagcagcgacct
 35 ttaagcagttagaagatttaagaactaataaagaacagctttgggatatgctaataaa
 tttaaacgtaaattacatccatgaaggtaagcagtatgagatgattatagatatatta
 gaaaatgatgcttcagaaatatattttgaattgcgtagctctcgatctgaatctacactg
 attacttataataaacacgaaaataaaactcactttagaacgtaccgatagtgggacacta
 ccatcaaattgctgatggaacaacgcgttctaccatttttagattcaccattaaaacagtta
 40 caaattttgtggatacatctagatcgaaatattctgtaattgatggtagcgtggtttta
 acctcacgtattttccaaatgaggatgctacaggtataaaagcttcgactgaatctggt
 caagtatatttaaaattcactaaatatgaattaaaagggtga

Sequence 784
 45 MIGDLNFNLFDFHESFQELDNGFDYAPQTFVDADGQRILIGWMGLPDTEYPTDKEGWA
 HCLTIPLVLTIEGKLRPFKQLEDLRTNKETALGYANKFKRKLHPYEGKQYEMIIDIL
 ENDASEIYFELRSSRSESTLITYNKHENKLTLERTDSGTLPSNVDGTRSTILDSPLKQL
 QIFVDTSIEIFCNDGERVLTSRIFPNEDATGIKASTESQVYLKFTKYELKG*

Sequence 785
 50 Contig_0515_pos_13289_12330,
 is similar to (with p-value 0.0e+00)
 >pir:pir|S20799|S20799 hypothetical protein 7 - Staphylococcus aureus
 >pir:pir|S58482|S58482 hypothetical protein 7 - Staphylococcus aureus
 55 >gp:gp|X52543|SAAGRAB_7 S.aureus agrA, agrB and hld genes. NID: g46505.
 atgagacgtctatttgctattggagaggcattaattgatttcataccaaatgtaacgcat
 tcaaaattaaagatgttgacaatttagtcgacaagttggtggcgaccgtgtaacgta
 gcggctacagtaagtaattagggtggaaaatcagaaatgataacacaactaggaatgac

15

MRRLFAIGEALIDFIPNVTHSKLKDVEQFSRQVGGAPCNVAATVSKLGGKSEMITQLGND
AFGDIIVETIEQLGVGTQYIKRTNKANTALAFVSLQDDGQRDFSFYRKPSADMLYQPENI
DDIQVFQDDILHFCSDVLIESDMKYAHEKMIIEKFESVDGTIVFDPNVRLPLWEDKLECQR
TINAFIPKAIKIVKIDSEELLFITGKRNEDEAIQSLFRGQNVVVIYTQGAQGATITYTKDDY
RIHHEGYQVQADITTDGAGDAFAGIIYCILESRHSECKDLFKEKGKDILAFSNRVAALT
TKHGAIESLPTKEDIKYN*

25 Contig_0515_pos_7564_6104,

atgcttacgggctttgctttcatggtaactacatcattattcagtcaccaagcacatgct
gaaggtaatcatcctattgacattaatttttctaaagatcagattgatagaaatcacagct
aagagcaatattatcaatcgagtgaaatgacactagtcgcacaggaattagtatgaattcg
30 gataatgatttagatacagatatcgtttcaaatagtgactcagaaaaatgacacatatta
gatagtgttcagactcagatagtgacttagattcagatagtgacttagcattcagacag
gactcagattcagatagtgactcagattcagatagtgactcagattcagacagtgattca
gactcagatagtgactcagattcagacagtgattcagactcagatagtgattcagattca
gatagtgttcagattcagacagtgactcagactcagacagtgattcagattcagatagtg
35 gattcagattcagatagtgattcagattcagatagtgattcagattcagacagtgactca
gactcagacagtgattcagattcagacagtgactcagattcagatagtgactcagattca
gatagtgttcagactctggtacaagttcaggtaagggttcacataccggaaaaaacct
ggtaacctaaaggaaatacaaatagacctctctcaagacatacgaattcaaccccaaagg
cctaaatacaatcaaaccaatcaaaacaatatataaacaatataaaccatatataaaccta
40 ataatcatcacgtactagtggagatagtgcgccttttaaacgtcaacaaaatattatt
aattctaacttaggtcatagaaatcaaaataatataaatcaatttatatggaacaaaat
ggcttttttaaatctcaaaataataaccgaacatagattgaatagtagtgataataccaat
tcattaattagcagattcagacaatttagccacgggtgcttataagtaacaatccggttttg
attaatcaagtaaaaaatttgaattcaattagatggaagggtgacagatagtgacattttg
45 agcttggttaggaagcaatcatttagaggaaatgaatttttaattcattacaaaaaggg
acaagctattttcagatttcaatattttaatccacttaattctagtaaatactatgaaaat
ttagatgatcaggttttagctttaattacaggagaaatcggtcaatgccagaacttaaa
aaacctacggataaagaagataaaaaatcatagcgtcttcaaaaaccatagtgacagatgag
ataacaacaaaataatgatggacactccaagattatgataagaaaaagaaaatacatcga
50 agtcttttattcgtttaagtattgcaataattggaatttttctaggagtcactggactatat
atctttagaagaaaaagtaaa

[illegible]

SLFRKQSFGRNEYLNSLQKGTSYFRFQYFNPLNSSKYYENLDDQVLALITGEIGSMPELK
KPTDKEDKNHSVFKNHSAD EITTNNDGHSKDYDKKKKIHRSLLSLSIAIIGIFLGVGTGLY
IFRRKE*

5 Sequence 789

Contig_0515_pos_4789_3170,

is similar to (with p-value 0.0e+00)

>gp:gp|U13618|SEU13618_2 Staphylococcus epidermidis 9759 heat shock protein 10 (hsp10) and heat shock protein 60 (hsp60)

10 genes, complete cds. NID: g535340.

atggcaaaagatcttaaattctctgaagatgcgcgtcaagcaatgttacgtggtgttgat
aaattagcaaacgctgtaaagggttacaattggacctaaaggcgaaatgtggttctagat
aaggattacacaacacctttaattaccaacgatggtgtaacaattgctaaggaaatagag
ttagaagatccatagtgaatattgggtgcaaaattagtgcaggaagttgcgaataaaaca
15 aatgaaatcgctggggacggtacaactacagcaacagttttagcacaatcaatgattcag
gaaggctctaagaatgttacaagtggtgcaaatcctgtaggcttaagacaaggtattgac
aaagcagtgcaagtggtatagaagcgcttcagagatttctcaaaagggtgaaaataag
aacgagatagcgcaagttggagctatttcagcagcagatgaagaaatcggtcgctacatt
tctgaagcaatggataaagtaggttaacgatggcggttatcactattgaagaatcaaatggg
20 ttaatacagaattagaagtagttgaaggatgcaatttgatcgcggttatcaatcacca
tatatggtaactgactcagataaaatgatagctgaattagaacgtccatatatattagta
acggataagaaaatttcatcattccaagatattcttcattattagaacaagttgtgcag
gctagtcgaccaattttaattggtgcggtgaagtagaaggcgatgcacttactaatatt
gttttaaaccgtatgcgtggaacatttactgctgtagcagttaaagccccaggatttggt
25 gatcgacgtaagcaatgttagaagacctagcaatattaaactggtgctcaagtcattact
gatgatttaggtttagaacttaagatgcattctctgatagctaggtagctgtaataaa
gttgaagtgactaaagatcatacaacagtcgtagatggtaatggtgatgaaaataatatt
gatgctcggtgtaggtcaaattaaagcacaattgaagaaactgattcagagtttgataaa
gaaaaattacaggaacggtttggcaaaactagctggcggtgtagctgttatcaaagtaggg
30 gctgcaagtgaaacagagcttaagaacgtaaaatgaagattgaagacgcattaaattca
acacgtgcgggcggtggaagaaggtatcggtgctggtggtggtactgcttagtcaatata
tatcaaaaagtaagtgaattaaagcagaaggtgatgttgaaacgggtgttaataatcgta
ttaaagcattacaagcacctgttagacaaattgctgaaaatgcaggattagaggggtca
attattgttgaaacgtttaaaacatgctgaagcgggctgtggtttcaatgcagcaacaaat
35 gaatgggttaatatgttagaagaaggtatagtagatccaactaaagtaactcgttcagcg
ttacaacatgcagcaagtgtagctgctatgttcttaacaactgaagcagtcgttgctagt
attccagagccagaaaataatgaacaacctggaatgggtggcatgccaggtatgatgtaa

40 Sequence 790

MAKDLKFSEDARQAMLRGVDKLANAVKVTIGPKGRNVLDKDYTTPLITNDGVITIAKEIE
LEDPYENMGAKLVQEVANKTNEIAGDGT TATVLAQSMIQEGLKNVTSGANPVGLRQID
KAVQVAIEALHEISQKVENKNEIAQVGAISAADEEIGRYISEAMDKVGNDGVITIEESNG
FNTELEVVEGMQFDRGYQSPYMTDSDKMIALERPYILVTDKKISSFQDILPLLEQVVQ
45 ASRPILIVADEVEGDALTNIVLNRMRGTF TAVAVKAPGFGDRRKAMLEDLAILTGAQVIT
DDLGLELKDASLDMLGTANKVEVTKDHTTVVDGNGDENNIDARVGQIKAQIEETDSEFDK
EKLQERLAKLAGGVAVIKVGAASETELKERKLRIEDALNSTRAAVEEGIVAGGGTALVNI
YQKVSEIKAEGDVETGVNIVLKALQAPVRQIAENAGLEGSIIVERLKHAEAGVGFNAATN
EWNMLEEGIVDPTKVTRSALQHAASVAAMFLTTEAVVASIPEPENNEQPGMGMPGMM*

50

Sequence 791

Contig_0515_pos_2752_1793,

putative peptide of unknown function

55 atgaaagacaacaaacctaataattcgaaattaattcaaacatattttaagtaagaaaact
ttaagatattggtacagcaagtgcatataacattggcactctattttatttaacagtaacgta
actgtgtatgctggtgaaaataactgcaaaccaaaatcaaggaacatcaccaaaaacttca
cagacagcacctacaaataataactgaaaatacagatgccacagccataacaacagatcaa
aataataatgatgaagaagaatacagatgcgtcatatgaacttccaattctttatgtaact

gtctggctagatgatcaaggaaatattattaaagatgctgtggaagatgctaaaaccct
 gcttcagaaaggcaaccggtgaaaattcctgggtaccaacattatagaacttctgtgagt
 gacggaattactaagttttatctcgtaaaattagcactgcacaatcacctatagttgaa
 aataatcaacaagataataatacaaaataaagttgttgaaacaaccaatcaaaataaagat
 5 gaagtgaatggaaaagaacaaaatcaagcaatacttcagtaacaaatacacaaattacc
 aaaaacgagaaagacgaagacacaaaaacactaaagaaagataaaagacgagaaagaatct
 aaagacacaaaaacaccaaagaaagacaaaagaaagacataaaaaactccgaagaaa
 gatagagaagagaaaaaaccagtaataccaaaaagcggcaagacgagaaagacacaaaa
 ataactaagaaagacaaagaagacgaaattacaacaacttccaagaaagataataacaat
 10 gatgtacaagataaattaccggaacaggtaaaaacacgatattcaaaatcctgcttta
 ataattgttactgtgctggttttaggtttatttaggattatttagaataaaaataagagaatag

Sequence 792

15 MKDNKPNNSKLIQTYLSKTLRYGTASALTLLALYLNSNVTYADENTANQNQGTSPKTS
 QTAPTNNNTENTDATAITTDQNNNDEEEYDASYELPILYVTVWLDDQGNIIKDAVEDAKTP
 ASERQPVKIPGYQHYRTSVSDGITKFIYRKISTAQSPIVENNQDNNNTNKVVETTNQNKD
 EVNGKEQNQANTSVTNTQITKNEKDEDTKTLKKDKDEKESKDTKTPKKDKKKDIKTPKK
 DREEKKPVIPKSGKDEKDKITKKDKEDEITTSKKDNNNDVQDKLPETGKTNDIQNPAL
 20 IMLLAGLGLLGLFRNKIRE*

Sequence 793

Contig_0517_pos_750_1070,
 is similar to (with p-value 5.0e-30)
 25 >sp:sp|P54453|YQEH_BACSU HYPOTHETICAL 41.0 KD PROTEIN IN NUC
 B-AROD INTERGENIC REGION. >gp:gp|D84432|BACJH642_92 Bacillus
 subtilis DNA, 283 Kb region containing skin element. NID: g
 2627063. >gp:gp|Z99117|BSUB0014_47 Bacillus subtilis complet
 e genome (section 14 of 21): from 2599451 to 2812870. NID: g
 30 2634966.
 atgatacctggtgtatcaaacataaatgatttttcgtctaatggaatatctatcatatct
 aaagttgttcctggaaagcgtgatgtagttactacatctttttctcccacactctgttca
 attaatattattaatgaatgtagattttccaacattcgttgtaacctacaatgtatacgtca
 tctttatttcttacatggtttatagattgcaataattcatcaatccccaacctttattt
 35 gcagaaataagaacgacatcttctgcttctaattccatattttacgagcagattttctcaac
 cattcttttacacgtcgatga

Sequence 794

MIPGVSNINDFSSNGISIIISKVVPGRDVTTSFSPTLCSINLLINVDFTFVVPTMYTS
 40 SLFLTWFIDCNSSIPQLFAEIRTTSSASNPLYRADFLNHSFTRR*

Sequence 795

Contig_0517_pos_2983_3453,
 is similar to (with p-value 4.0e-29)
 45 >gp:gp|D50453|D50453_106 Bacillus subtilis DNA for 25-36 deg
 ree region containing the amyE-srfA region, complete cds. NI
 D: g1805369.
 atgagaattgcacctaattgtgattggtagaatccaaccattaatcgaccagctattata
 agtaaacacaccggtttaccaataaataagaaaacaaaagttgaaattacaataaatgta
 50 ataacgataagattatttttattgagtaacgatttgtgtagtgttttaaaaatgttgcg
 cttgtatatgcagaaccaattactgaggacattgctgctgcaaatattactacgcaaaa
 atatttttacctataggacctaattgcatgttggaactgatgctgggtggattttctgaa
 ctaagcgtaacgccagttacaacaacacctagtacagctaaaaacaataaggtgcgcatg
 acaccagttgttaaaatcctgctacagcagatcgatttacgaaaggaaggtatgactta
 55 cttttataccagaatctagaattctatgtgcacctgcaaaagtaataataa

Sequence 796

MRIAPNVIGRIQPLIAPAIISKLTGLPINKKTKVEITINVITIRLFLLSNDLCSVFKNVA
 LVYAEPITEDIAAANITTPKIFLPIGNACWKTDAGGFSELSVTPVTTTPSTAKNNKVRM

TPVVKIPATADRFKGRYDLFPFIPESRILCAPAKVI*

Sequence 797

Contig_0517_pos_8261_7662,

5 is similar to (with p-value 3.0e-41)
 >sp:sp|P42967|YCSJ_BACSU HYPOTHETICAL 63.8 KD PROTEIN IN SIP
 U-PBPC INTERGENIC REGION. >gp:gp|D38161|BAC39R_12 Bacillus s
 ubtilis genome around 39 degrees region encoding 17 ORFs, co
 mplete cds. NID: g1032472. >gp:gp|Z99106|BSUB0003_56 Bacillu
 10 s subtilis complete genome (section 3 of 21): from 402751 to
 611850. NID: g2632653. >gp:gp|D50453|D50453_108 Bacillus su
 btilis DNA for 25-36 degree region containing the amyE-srfA
 region, complete cds. NID: g1805369.
 atgataccttcataatcgctgctatttctaatttacttttgataaatcggggataaacggaact
 15 gaattattagaaaatttagaactcgatgaaaattcgaacagtagaaaacaatcacatttt
 aaacagcgtatcattcatatacctgtattatatgggtggagattttggtccagatttatca
 gaagtagctaacgttaataaattaagtcaggaagaagttattcaaatacacatacacaaca
 ccttatctaattctatatgcttgggtttatgccgggtttccatacttaggtggattggat
 gctaagttgcacacacctagacgggtctgaacctagaatcaaaattaacgctggttctggt
 20 ggaatagcaataatcaaacaggttttatcttatgactcacctggtggttggcagata
 attggtgcgcacaccaataaaagtcctttgatttaaataggacaccaatgacgttatatgaa
 gctggtgattacatacaattttatagtataaattatcaagagtttgaaaaaatatcaaac
 gatattaataaaggaaaatttgatatagataagtggtgacatatcaagatgagtattaa

25

Sequence 798

MIPSYRAILYFDKSGINGTELLENLELDENSNSRKQSHFKQRIIHIPVLYGGDFGPDLS
 EVANVNKLSQEEVIQIHTQQPYLIYMLGFMPGFPYLGGLDAKLHTPRRSEPRIKINAGSV
 GIANNQTGLYPMDSFGGWQIIIGRTPIKVFDLNRTPTMLYEAGDYIQFYSINYQEFKISN
 30 DINKGKFIDIKWVTYQDEY*

Sequence 799

Contig_0517_pos_7306_6668,

35 is similar to (with p-value 2.0e-25)
 >sp:sp|P44298|YBGK_HAEIN HYPOTHETICAL PROTEIN HI1730. >pir:p
 ir|B64041|B64041 hypothetical protein HI1730 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32845|U32845_11 Haemophil
 us influenzae Rd section 160 of 163 of the complete genome.
 NID: g3212236.
 40 gtggctcaaaagctattctacacatgtagaagtggaatggcggtatttaaaggctcgtgca
 ttaaagaaatacagatgttattgcaactcaagtaaatcataactataaaaactaatttagga
 aaaacaattgattttctcatctatacctgataataattacatacacgctcattgaaggacct
 cagatcaatgaatttgatgaagaaacgatagctaaattcgtaaatagcgatttcaaaatt
 tctgatcaatcagatcgaatgggatacagattaaaaggtaatacagtagccctaaaaat
 45 agtgctgatattcttctgaacctgtcgctttgggaagtattcaagtacctaacgatggt
 aatcccattattcttttaaatgataagcaacaattggtggttatacaaaaattgcaacg
 gtaacacaattagatttaagaaaatttagcacagatgaagcctggagacattatacagttt
 aaatggataactgttgagaagcttcaaaaaagcttaaaagaatttaataactaaatttgaa
 caattattaaagcggttttgatgagcaaccattgtttaacctaaatcaacttagacatact
 50 tctaataaaatcgcaaaaataattaaggagtagataa

Sequence 800

55 VAQSYSTHVRSGMGGFKGRALKKYDVIATQVNNHYKTNLGKTI DFSSIPDNNYIHVIEGP
 QINEFDEETIAKFVNSDFKISDQSDRMGYRLKGNTPPKNSADIISEPVALGSIQVPNDG
 NPIILLNDKQTIGGYTKIATVTQLDLRKL AQMKPGDIIQFKWITVEEASKKLKEFN TKFE
 QLLKRFEQPLFNLNQLRHTSNKIAEIIKEDR*

Sequence 801

Contig_0517_pos_6667_6227,

putative peptide of unknown function

atggatattaaaaaattgaggaagtcattaaattggtaaaagctaataatgatgtaaaaaaa
 ttttaagtataaggactctcataatgaaatagaacttgattttactaatggagcatctcaa
 caacattcgcaacaatcatctcaagatattcaacaagagaatattaaatccttagatgaa
 5 aagcaagagtccatatcaaataaccagcaagagattaaatctcctatgggttggaacattt
 ttcttacaagatagtaaagaactaacagaacctaaagattaaagttggcgatactgtaact
 gaaggagatattatcggttacattgaagctatgaaagttatgaatgaagtaactacggat
 gttactggtagagtcactgaaatattagtagaacatggagacaatggtgaatatgatcag
 ctactagtcagaggttaaataag

Sequence 802

MDIKKIEEVIKLVKANDVKKFKYKDSHNEIELDFTNGASQQHSQQSSQDIQQENIKSLDE
 KQESISNDQOEIKSPMVGTFFLQDSKELTEPKIKVGDVTEGDIIGYIEAMKVMNEVTTD
 VTGEVTEILVEHGDNVEYDQLLRVK*

Sequence 803

Contig_0517_pos_6216_4855,

is similar to (with p-value 0.0e+00)

>sp:sp|P49787|ACCC_BACSU BIOTIN CARBOXYLASE (EC 6.3.4.14) (A
 20 SUBUNIT OF ACETYL-COA CARBOXYLASE (EC 6.4.1.2)) (ACC). >gp:
 gp|U36245|BSU36245_2 Bacillus subtilis biotin carboxyl carri
 er protein (accB) and biotin carboxylase (accC) genes, compl
 ete cds. NID: gl055244.

atgatgtatcgatgtttgattgcaaatagaggcgaaatagcagtaagaattataagagct
 25 tgtagagagcttaacatagagacagttgccatttatgcaaaagtgatgaaaatagctta
 catgtaagtttagccgatcaagcaatatgcataggtgaagcaaatccattagacagttat
 ttaaatattgatcgatattatatctgctgcaaaagttacagaatcaaacgtaattcaccct
 ggctatgggtttcttatcggaatctacgaattttgcgaaagccggtgaagacaatcatata
 cattttattggacctagtaagacaactatggaaatgatgggggataaaattactgccaga
 30 caaactgttaaaacagcaggagtagctgttataccaggttctaattgatgctgttcaaagt
 gtagatgaaattaaattattatccaaagaaataggatttccagttgtactaaaagcagct
 agtgggtgggtgggaaagcatcagaattgttaaagaagcatctcatttggatcaggct
 ttgaaagaagctaaaagtgaaggacaaaatattttaattgatgatcgagtgtagtagag
 gcggttcataccagtagcaaaacatgtagaagtgtagattatcgagagcggtaaaaataac
 35 tatgttcacttaggtgaacgcgattgttctgttcaacgaaagaatcaaaaattaatagaa
 gaagcgccttgtgctgcattactgaagaagaagaacaagaatatgtggcgacgcagtt
 aaagtagctcaagcttcaagatatcgtagtgctggaacaatagaatttttagttacagaa
 gatgcacattattttattgaaatgaatgctcgtattcaagttgaacatacagttacagaa
 atgctgctgatagagacctattacaagctcagttatatatttataacacacgggtgaatta
 40 ccattcactcagaaagatattttatttaattggtcatgtaattgaggcgcgtataaatgct
 gaaaatcctgaaaaaaactttttaccactccaggaaaagttataaaattacacttacca
 caaggatttaatatagctgtagattctttactttacacaggttatcagggtttctccttat
 tatgattcacttgtagctaaagtgattgttaaaggattctaataagacaaactgctattaat
 aaattaaaagttgctgtagatgaaatggtcatcgaaggttttactactacagctgacttt
 45 ttatatgcgggttttaattatccaatatatgcaaaaggcgatgccagtaaaagtagatata
 aaatttcttgaaaaacatcaaatacattaaagaggtgaaatga

Sequence 804

MMYRCLIANRGEIAVRIIRACRELNIETVAIYAKGDENSLHVSLADQAICIGEANPLDSY
 50 LNIDRIISAQVTESNVIHPGYGLSESTNFAKAVEDNHIHFIGPSKTTMEMMGDKITAR
 QTVKQAGVPVPGSNDVQSVDEIKLLSKEIGFPVVLKAASGGGGKIRIVKEASHLDQA
 LKEAKSEGQKYFNDDRIVYEAFIPVAKHVEVQIIGDGKNNYVHLGERDCSVQRKNQKLIE
 EAPCAALTEERRTRICGDAVKVAQASRYRSAGTIEFLVTEDAHYFIEMNARIQVEHTVTE
 MRADRDLLQAQLYLLTHGELPFTQKDILFNHGVIEARINAENPEKNFLPTPGKVNKLHLP
 55 QGFNIRVDSLTYGYQVSPYYDSLVAKVIVKDSNRQTAINKLKVALDEMVIIEGFTTTADF
 LYAVLNYPIYAKGDASKVDIKFLEKHQIIIEVK*

Sequence 805

Contig_0517_pos_4088_2844,

is similar to (with p-value 2.0e-62)

>pir:pir|G64138|G64138 branched chain aa transport system II carrier protein (braB) homolog - Haemophilus influenzae (strain Rd KW20) >gp:gp|U32845|U32845_9 Haemophilus influenzae Rd section 160 of 163 of the complete genome. NID: g3212236.
 5 atgggggaaaatacaaaaacaagatttcaatcaaaaaggacaaaattttaattcacaaaa
 aaacatagacgattattatatggttcagttttttaatggctacatcagctattggtcca
 gcatttctgactcaaactgcagtggttactgcacaattttatgctagttttgcatttgca
 atattaatttctattattatagatataggcgctcaaataaatatttggagaatattagt
 10 gtaactgggattacgtggacaagaaatatctaataaagtattacctggacttggactatt
 atctccatactaattgcatttgggtggtctcgcatttaacataggtaatattgctggtgca
 gggttaggttttaaatgcaatggttggcttgatgtaaaatgggtgctgcaataacagct
 atttttgcgatacttattcttggtagtagaagtggtcagaaaaataatggatgttattagt
 atgattctaggtatcgtaatgatttttagtagtgcgttatgtcatgggtgttcaaattccc
 15 ccttatggagatgcattagtagacatacatttgcacctgaacatcctttcaaacttatatta
 cctataaattacattagttggtggtacagtaggggttatattacttttgcaggtagcat
 agaattctagattctggtataaaaggtaagtcataccttcttctcgtaaatcgatctgct
 gtagcaggtattttaacaactggtgtcatgcgcaccttattgttttagctgtactaggt
 gttgttgtaactggcggttacgcttagttcagaaaaatccaccagcatcagttttccaacat
 20 gcattaggtcctataggtaaaaatatttttggcgtagtaaatatttgcagcagcaatgtcc
 tcagtaattggttctgcatatacaagcgcaacatttttaaaaacactacacaaatcgta
 ctcaataaaaataatcttatcggtattacatttattgtaatttcaacttttgtttctta
 tttattggtaaaccggtgagtttacttataatagctggtgcgattaatggttgattcta
 ccaatcacattaggtgcaattctcattgcaagtaggaaaaatctatcggttgtaattac
 25 caacaccaacatggatgcttgttttgggtattatagccgtaattgtcacataatgact
 ggtatcttttcattacaagatttagcaagttctttggaaggttaa

Sequence 806

MGENTKQDFNQKGQNFKFTKKHRRLLYGSVFLMATSAIGPAFLTQTAVFTAQFYASFAPA
 30 ILISIIIDIGAQINWRILVVTGLRGQEI SNKVL PGLGTII SILIAFGGLAFNIGNIAGA
 GLGLNAMFGLDVKGAAITAIFAILIFVSRSGQKIMDVISMILGIVMILVVAYVMVSNP
 PYGDALVHTFAPEHPFKLILPIITLVGGTVGGYITFAGAHRLDSGKIGKSYLPFVNRSR
 VAGILTTGVMRTLLFLAVLGVVVTGVTLSSENPPASVFQHALGPICKNI FGVIIFAAAMS
 SVIGSAYTSATFLKTLHKSLNKNLIVITFIVISTFVFLFIGKPVSLIIAGANGWIL
 35 PITLGAILIASRKKSIVGNYQHPTWMLVFGIIAVIVTMTGIFSLQDLASLWKG*

Sequence 807

Contig_0517_pos_1915_1412,
 is similar to (with p-value 4.0e-46)
 40 >sp:sp|P54452|YQEG_BACSU HYPOTHETICAL 20.1 KD PROTEIN IN NUC
 B-AROD INTERGENIC REGION. >gp:gp|D84432|BACJH642_91 Bacillus
 subtilis DNA, 283 Kb region containing skin element. NID: g
 2627063. >gp:gp|Z99117|BSUB0014_48 Bacillus subtilis complet
 e genome (section 14 of 21): from 2599451 to 2812870. NID: g
 45 2634966.
 atgccaaatgcatatgtgaaatcaatatttgaaattgatatagaaaaacttgccgatagt
 ggtgttaaaggtatcataactgatttagataatacacttgttggttggtggttaaagaa
 cctactaagggtgttaaactcatggttgcgaaggctaaagatttaggaataactgtcaca
 attgtgtcaataataataaaagtgcagtagtcaagtttctcaagtaatttaggtgtagat
 50 tatataattcaaagcacgtaaaccgatggggaagcctttaagatggctattaaaaaatg
 aaaattcaaccgagagaaaccgttgtttagtaggatcaaatgcttactgatgtgttgggt
 ggcaattgtaattggtttatatacaattatggtagtagctgttaaaccgactgatggatta
 attacaaagttaaactcgattaattgaaagacgattattaaatcatttttagaaaaaagggt
 tatattaaatgggaggaaaattga
 55

Sequence 808

MPNAYVKSIFEIDIEKLADSGVKGIITDLDNTLVGWDVKEPTKGVKSWFAKAKDLGITVT
 IVSNNNKSRSVSSFNGLVDYIFKARKPMGKAFKMAIKMKIQPRETVVVGDMQLTDVFG
 GNCNGLYTIVMVPVKRTDGLITKFNRLIERLLNHFRKKGYIKWEEN*

- Sequence 809
Contig_0517_pos_805_311,
is similar to (with p-value 6.0e-39)
- 5 >sp:sp|P54453|YQEH_BACSU HYPOTHETICAL 41.0 KD PROTEIN IN NUC
B-AROD INTERGENIC REGION. >gp:gp|D84432|BACJH642_92 Bacillus
subtilis DNA, 283 Kb region containing skin element. NID: g
2627063. >gp:gp|Z99117|BSUB0014_47 Bacillus subtilis complet
e genome (section 14 of 21): from 2599451 to 2812870. NID: g
10 2634966.
atgatagatattccattagacgaaaaatcatttatgtttgatacaccaggtatcattcaa
tcacatcaaatgacaaattatgtatatgaaaatgagttgaaaatcattatacctaaaaat
gaaataaagcaacgtgtgtatcaacttaatagaaaaacagacattattttcggaggattg
gcacgcattgattatgtatctggtggtgtaaagaccacttggttcttttcaaagat
15 ttaaattattcatagaactaaaaccgagaaagctaatagtttatggaaatcccaattaggc
gcattgctttcaccgcctcaagatgcacaacaatttaattcttaataatgataaagcagta
agactggaaactggtaaaactaaacgtgacatcatgatatctggttaggattcataact
attgatgctggtgcaaaagtgatagttcgtgttccaaacatgtagatgttattttaaga
aattcaattctttaa
- 20 Sequence 810
MIDIPLDEKSFMDTPGIIQSHQMTNYVYENELKIIIPKNEIKQRVYQLNEKQTLFFGGL
ARIDYVSGGKRPLVCFNSDLNIHRTKTEKANDLWKSQGLALLSPQDAQQFNLDVKAV
RLETGKTKRDMISGLGFITIDAGAKVIVRVPKHVDVILRNSIL*
- 25 Sequence 811
Contig_0517_pos_0_304,
putative peptide of unknown function
gtgataaaagtgaattttgcagtaattgaaacccccatttctcattcattatcgccattg
30 atgcatcatgctaattttcaatctttaaatttgaaaacacgtatgaagcgataaatgta
ccagttaatcaattttcaagacattaaaaaataatttcagaaaagagtattgatggattc
aatgttactattccacataaagaacgtattattccgtacctagatgatattaatgaacaa
cgaaatctgttgggcggttaatacagtttttagttaagatggttaagtggattggttat
aata
- 35 Sequence 812
VIKVKFAVIGNPISHSLSPMHANFQSLNLENTYEAINVVPVNFQDIKKIISEKSIDGF
NVTIPHKERIIPLYDDINEQAKSVGAVNTVLVKDGKWIGYNX
- 40 Sequence 813
Contig_0518_pos_2682_1942,
is similar to (with p-value 1.0e-70)
>sp:sp|P32816|GLDA_BACST GLYCEROL DEHYDROGENASE (EC 1.1.1.6)
(GLDH). >pir:pir|JQ1474|JQ1474 glycerol dehydrogenase (EC 1
45 .1.1.6) - Bacillus stearothermophilus >gp:gp|M65289|BACGLDA_
2 Bacillus stearothermophilus glycerol dehydrogenase (propos
ed gld) gene, complete cds. NID: gl42976.
atggatgaccaacagcagcagtatctgttatttataacgaagatggatcatttagtggt
tatgaattctaccctaaaaaccctgatacagttatcgtagattctgaaattgttgcaaa
50 gcacctgtacgttttatttgcacaggtatgagtgatggttttagcaacattaatcgaagt
gaatctacacttcgtagacaagggcaaacatgttccatggcaaacctacattagcaagt
ttagcaatcgctcaaaaatgtgaagagggtatttttgaatatggttacagtgccttact
tctgtagaaaaacatatcgtgacaccacaagtagatgctgtgattgaagccaatacatta
cttccaggttttaggatttgaaaacggcgatttagcaggtgcacacgcaattcataatgga
55 ttcacagcttttagagggatattccaccacttaactcatggtgaaaaagtggcatacggg
attttagtacaatttagtacttgaaaatgcgccaaactgaaaaattcatgaaatacaaaaca
ttcttcgataatatcaatatgccaacaacattagaagggtcttcacattgaaaacacaagt
tatgaagaatttagttcaagtaggtgaacgtgcattaacaccaaatagatacgtttgctaac
ttaagtataaaatcactgctgatgaaatcgacagcgaattttaactgttaattgattta

tctaaaagtcagttcaactaa

Sequence 814

MDAPTAAVSVIYNEDGSFSGYEFYPKNPDTVIVDSEIVAQAPVRLFASGMSDGLATLIEV
 5 ESTLRRQGNMFHKGPTLASLAIQKCEEVIFEYGYSAITSVEKHIVTPQVDAVIEANTL
 LSLGLGFENGLAGAHAIHNGFTALEGDIHHLTHGEKVAYGILVQLVLENAPTEKFMKYKT
 FFDNINMPTTLEGLHIENTS YEELVQVGERALT PNDTFANLSDKITADEIADAILTVNDL
 SKSQFN*

10 Sequence 815

Contig_0518_pos_909_334,

putative peptide of unknown function

atgaatgtagcagatatcaaagcacgcttattagatttagaaaaatacttttaagaaaaa
 gaaagtgaactgactgatttagacagagctatcggtgatggagatcatggtgtaaatatg
 15 gtcagagggtttcgaaactttaaaagaaaaaatagatgatcaaagtatgcaaagcgctattt
 aaatcaacaggtatgacattaatgtctaacgtagggtggtgcttctggaccattatacggg
 tttggttttatcaaaatggcgagtgacgtgaatgatgaaattgatcatgataatcttaaa
 gaggtacttaaaagcgtttgctgatggcattcaacaacgtggtaaaagtcgaattaaatgaa
 aaaacgatgtatgatgttatcgaaagtcgcagagagaagctgttgaaaaaaatgaaacagta
 20 gatctagataaaactacaatcatttgctaataaaccagatagtgtagctactaaaggc
 cgtgcatacatatttaacgaagcttcaaaaggttatattgatcctggtgcacaaagtagt
 gtttatattcttaatgcaattataggaggagagtaa

Sequence 816

25 MNVADIKARLLDLENTFKEKESELTDLDRAIGDGDHGVNMVRGFEHLKEKIDDSMQALF
 KSTGMTLMSNVGGASGPLYGFYFIKMASAVNDEIDHDNLKEVLKAFADGIQQRGKVELNE
 KTYDVIERAREAVEKNETVDLDKLSFANETKDMVATKGRASYFNEASKGYIDPGAQSS
 VYILNAIIGGE*

30 Sequence 817

Contig_0518_pos_0_330,

putative peptide of unknown function

atgacatctatagtagtagtaagtcataagtcataaaatcgagaaggtgttaaacaaatta
 atcaatcaaatgactgacggtggtgttgaccttattgccgttggtggcttaagtgcagat
 35 gaaatcggtacatcatttgatcaaatcgctctctgttaattgaacttgaaaatgatgcg
 ctttggttctatgacatcggttcagcaggtatgaatttagacacagctttgaaatgtac
 gaaggtgaccacaaaattgttaaaatggaagcgccaatcggtgaaggaagctttattgca
 agtgttaggaattaaatcaaatatgagtata

40 Sequence 818

MTSIVVSVSHSHKIAEGVKQLINQMTDGGVDLIAVGGLSDDIEIGTSFDQIVSVINGLEND
 LCFYDIGSAGMNLDTALEMYEGDHKIVKMEAPIVEGSFIASVGIKSNMSI

Sequence 819

45 Contig_0519_pos_4834_5202,

is similar to (with p-value 8.0e-23)

>gp:gp|AF026147|AF026147_6 Bacillus subtilis YojA (yojA), Yo
 jB (yojB), YojC (yojC), YojD (yojD), YojE (yojE), YojF (yojF
), YojG (yojG), YojH (yojH), YojI (yojI), YojJ (yojJ), YojK
 50 (yojK), YojL (yojL), YojM (yojM), YojN (yojN), and YojO (yoj
 O) genes, complete cds; and OdhA (odhA) gene, partial cds. N
 ID: g3169316. >gp:gp|Z99114|BSUB0011_110 Bacillus subtilis c
 omplete genome (section 11 of 21): from 2000171 to 2207900.
 NID: g2634230.

55 gtgaacgtgttggaaccaattaaagaacaagaagtgcctagatttattaacttcttactca
 aatcagcctgtttacctacacgttgaaacaacaaatgggtgcttatgcaaatcatttcgat
 caacgcgtattttaacgctggaacatttttaagaaatattgtcgtgacttttgaacatgca
 caacttaaaaggcggcgacaaagatccatatcgtgttaggtcttaaaattaaagatggtggc
 tgggtttacgtgcaaggacttacgcactatgaagttaatgagaataacgaatttttaatt

gcagggttttaattatgaaggacaattggctgctacaatagaataagtaaacagccattt
actatataa

Sequence 820

- 5 VNVLEPIKEQEVLDDLTSYSNQPVYLHVETTINGAYANHFDQRVFNAGTFLRNIVVTFEHA
QLKGGDKDPYRVGLKLKDGWVYVQGLTHYEVNENNEFLIAGFNYEGQLAATIEISKQPF
TI*

Sequence 821

- 10 Contig_0519_pos_5218_5883,
is similar to (with p-value 8.0e-26)
>gp:gp|AF026147|AF026147.7 Bacillus subtilis YojA (yojA), YojB (yojB), YojC (yojC), YojD (yojD), YojE (yojE), YojF (yojF), YojG (yojG), YojH (yojH), YojI (yojI), YojJ (yojJ), YojK (yojK), YojL (yojL), YojM (yojM), YojN (yojN), and YojO (yojO) genes, complete cds; and OdhA (odhA) gene, partial cds. N
15 ID: g3169316. >gp:gp|299114|BSUB0011_109 Bacillus subtilis complete genome (section 11 of 21): from 2000171 to 2207900.
NID: g2634230.
20 atgactgatgaaagacacgtacttgtgattttcccccatcctgatgatgaaactttttcg
tctgctggaactatcgcaagttatattgaaaaaggtattcccgtcacatatgcatgtctt
accctaggacaaaatgggacgtaattctaggaaccctccttttgcaacaagagaatcttta
ccatttatacgtgaacgtgagttagaagaagcatgcaagcaattgggattacagattta
aggaaaatggggttaagagataaaactgttgatttgaaccttacgatcaaatggatcaa
25 atgattcaatcacttattgacgaaacaaatccatcataattatttcgttctatcctaaa
tttgacgttcacccctgatcagcaggcaactgcagaagctgtagtacgtacagttggacgc
atgcataatcagatcgaccccgctttacactttagcggttagcaatgatgcacagaa
attcttgagaaacctgatattcaaaatgacatatctcaatatagtgatataaaacttaa
gcttttgaaagcacatgcttcacaaacaggaccatttttaaaacaacttgctagtcccgaa
30 atagatggtcaagcacaagtttcttaaaaatagagccattttggacatatcactttgaa
tcttaa

Sequence 822

- 35 MTDERHVLVIFPHPPDETSSAGTIAASYIEKGIPVITYACLTGQMGRNLGNPPFATRESL
PFIRERELEEACKAIGITDLRKMGLRDKTVEFEPYDQMDQMIQSLIDETNP SLIISFYPK
FAVHPDHEATAEAVVRTVGRMHESDRPRLTLVAFSNDASEILGEPDIQNDISQYSDIKLK
AFEAHASQTGPFLKQLASPEIDGQAQSFLKIEPFWTYHFES*

Sequence 823

- 40 Contig_0519_pos_3532_3086,
is similar to (with p-value 1.0e-29)
>sp:sp|P42405|YCKG_BACSU HYPOTHETICAL 19.0 KD PROTEIN IN TLP
C-SRFAA INTERGENIC REGION (ORF10). >gp:gp|D30762|BACYCK_10 B
acillus subtilis DNA around 28 degrees region of chromosome
45 containing yckA-H genes. NID: g710627. >gp:gp|D50453|D50453_49
Bacillus subtilis DNA for 25-36 degree region containing the amyE-srfA region, complete cds. NID: g1805369.
atggatgcagcagattacgaagtgcagcaagcagtaaaatatgggtgcagatattgttaca
attttaggtgtgtgctgaagatgcttcaattaaagcagcagttgaagaagcgcataaacat
50 ggaaaagcattgcttgttgatgatagcagtgcaaaacttagaacaacgtgctaaagaa
ctagatgagatgggtgcagactatatcgagttcatagagttacgacttacaagctgaa
ggaaaatctccattagacagcttgctgacagttaaatctgttatcaaaaactctaaggtt
gcagtagcaggtggtattaaaccagatactatcaaagatattgttgctgaagatccagat
ttagttattgttggtggcggtattgcaaatgctgacgatcctgtagaagcagcaaaacaa
55 tgtagagcagctattgaaggtaaataa

Sequence 824

MDAADYEVSQAVKYGADIVTILGVAEDASIKAAVEEAHKHGKALLVDMIAVQNLEQRAKE
LDEMADYIAVHTGYDLQAEKSPSLDSLRTVKSIVIKNSKVAVAGGIKPDTIKDIVAEDPD

LVIVGGGIANADDPVEAAKQCRAAIEGK*

Sequence 825

Contig_0519_pos_3084_2536,

5 is similar to (with p-value 7.0e-35)

>sp:sp|P42404|YCKF_BACSU HYPOTHETICAL 20.0 KD PROTEIN IN TLP
C-SRFAA INTERGENIC REGION (ORF9). >gp:gp|D30762|BACYCK_9 Bac
illus subtilis DNA around 28 degrees region of chromosome co
ntaining yckA-H genes. NID: g710627. >gp:gp|Z99105|BSUB0002_
10 174 Bacillus subtilis complete genome (section 2 of 21): fro
m 194651 to 415810. NID: g2632457. >gp:gp|D50453|D50453_48 B
acillus subtilis DNA for 25-36 degree region containing the
amyE-srfA region, complete cds. NID: g1805369.
atgagtgaatttaataattatcggtcttattcttgaagagttagattctactttatctcaa
15 gtagataatacagagtatgaacgttttgctaattgatgttataggtgcagatcgcatattt
acagctggttaaaggtcgttcaggttttggtgctaattggtttgcaatgcgcttaaatcaa
ttaggttaaaatgcctacgtttaggtgagtgcaacaacaccttcaattaaagaacatgat
ttgtttattattatttcaggttcaggttctacagaacatttaagattattagctgaaaaa
gcacaatctgaggggtgcaaaaattgtcttattaactacaaatgcggaatcgccaatcggt
20 aatcttgagagacgggtgttgaattgcctgcaggtactaaacatgatgttgagggttcg
aaacaaccacttggttagtttatttgaacaggcttcacttatattcttagatagtggtgta
ttacctttaatggatgcatttcacatttagtgaaaaacaatgcaagagaatcatgctaatt
tagaataa

25 Sequence 826

MSEFNRYRLILEELDSTLSQVDNTEYERFANDVIGADRIFTAGKGRSGFVANSFAMRLNQ
LGKNAYVVGESTTPSIKEHDLFIIISGSGSTEHLRLLAEKQSEGAKIVLLTTNAESPIG
NLAETVVELPAGTKHDVEGSKQPLGSLFEQASLIFLDSVVLPLMDAFHISEKTMQENHAN
LE*

30

Sequence 827

Contig_0519_pos_2414_1767,

putative peptide of unknown function

atgaattttgatagttatatatttttgattttgatggaacgctaattgatacaacaacatgt
35 cacgtcaaagctacgcaaagcgctttttaaagattaaatttagatgaacctacagaacaa
gctattttacatacatatcatttaaattttatataacaatttttaaagcgctagcttcacat
gaactgtctttttatcaaatagaaaaattaatagatgaatacaatcattgttttagcaac
gatgaaatacatcaatcaaaagaatataccggaataagtgaagcattaaaattttacat
aaccaaaagaaaaaaatattttagtgtctataaaagaaataactaacaactcaaaagtat
40 ttagattatctcggttaagccgttttataactgattcattaggtgtctgtattaaaaat
gaagacaaaacttctttgtgaaacgattcaaaatttgatacagaacatcatttaatgata
ggtaaaaccgtgtatataggggacacagcacagaatatcaagagtgcgaatcaagctcat
gtgcaaacatgcgctgtcacatggggagcacaatctgcacacgaattgttgcatgaaaaat
cctcattatattgttaatgatccagaagaatttttaacaattttataa

45

Sequence 828

MNFDYSIFDFDGLIDTTTCHVKATQSAFKRLNLDEPTEQAILHTYHLNLYNNFKALASH
ELSFYQIEKLIDEYNHCFSNDEIHQSKEYTGISEALKFLHNQKKKIFVVSNKEILTQKY
LDYLGSLSRFITDSLGVCIKNEDKLLCETIQNLQKHHLMIGKTVYIGDTAQNIKSANQAH
50 VQTCVWTGAQSAHELLHENPHYIVNDPEEFLTIL*

Sequence 829

Contig_0520_pos_483_1154,

55 is similar to (with p-value 9.0e-88)

>gp:gp|AF022796|AF022796_2 Staphylococcus carnosus molybdenu
m cofactor biosynthetic gene cluster, complete sequence. NID:
g3955197.
atgcctgatttaacgtccttttggtttcttttcgtgttgctttaatcagtacaatgata
gttactatattttggcatttttgatttctaaatggctatacaataaaaaaagatattgggta

aatctattagaaagttttatcattttaccaattgtgttaccacctactgtccttggtttt
 atactattaattatattttcaacaagaagtcctgtaggagaattctttactaatatctta
 cacttaccagttgtattttacattgacaggtgcagtgattgcatctgtcattgttagtttt
 ccccttatgtatcaacatacagtgatggttttcgaagtatagattcaaatggttaa
 5 actgcaagaacgatgggagcaagtgaacaaaaatatttcttaaattgggtgttaccatta
 tctaaacggttctattcttgcaggtattatgatgagctttgcaagagcaataggtgaattt
 ggtgctactttgatgggtgctggtatatacccagacaaaaacatacattgcctttagaa
 atttatttttttagtgagcaagggaaagaaaatgaagcatggttatgggtgcttgatta
 gttgcggttgcggttaactgtcatagcgaccataaatctggttaatcgtgatacgttttagg
 10 gaggttgattaa

Sequence 830

MPDLTSFWISFRVALISTMIVTIFGILISKWLYNKKRYWVNLLSFIIPLPPTVLGF
 ILLIIFSTRSPVGEFFTNILHLPVVFTLTGAVIASVIVSFPLMYQHTVNGFRSIDSMLN
 15 TARTMGASETKIFLKLVLPLSKRSILAGIMMSFARAIGEFGATLMVAGYIPDKTNTLPLE
 IYFLVEQKGKENEAWLWVLVLVAFVAVTVIATINLVNRDTFREVD*

Sequence 831

Contig_0520_pos_1155_1775,
 20 is similar to (with p-value 5.0e-73)
 >gp:gp|AF022796|AF022796_3 Staphylococcus carnosus molybdenu
 m cofactor biosynthetic gene cluster, complete sequence. NID
 : g3955197.
 atgctcacaaattaaagtgaatgggtgttcttaatcagacgaaaattaatataaaataaaag
 25 gatcaacaccctaagatatatgcgatacaggaccatctggaattggaaagacaacaatt
 ttaaataataattgccggtttgaaagctataaattattcatatataaagggttgtaaactg
 gtattaactgattcacgacaccatttgaatgttaagggttcaacaacgctcgtataggatat
 ctattttcaagattatcaacttttcccaatatgaatgtttataacaacataacgtttatg
 actaaaccttctgaacatatcaatgaacttattcatactctaaaaatagagcatttactt
 30 gaaaagtatcctgtgaccttatcaggaggtgaagctcagcgctcgcttttagcaagggcg
 ctaagtacgaaaccgatttgattttgcttgatgagccttttcaagtttagatgataaa
 acaaaaaacgaaggtatcaaattaattttaaaaatattcgaagcatggcaaattcctatt
 atatttgtaacgcattcaaattatgaagcgcaacaatggcgcatgagattataacaatt
 gaagattgtatacaaatatag
 35

Sequence 832

MLTIKVNGLVNLQTKININIKDQHPKIYAIQGPSGIGKTTILNIIAGLKAINYSYIKVGKR
 VLTDSRHHLNVKQQRRIGYLFQDYQLFPMNMVYNNITFMTKPSEHINELIHTLKEHLL
 EKYPVTLSGGEAQRVALARALSTKPDILLDEPFSSLDKTKNEGILKILKIFEAWQIPI
 40 IFVTHSNYEAQQMAHEIITIEDCIQI*

Sequence 833

Contig_0520_pos_1851_2852,
 is similar to (with p-value 0.0e+00)
 45 >gp:gp|AF022796|AF022796_4 Staphylococcus carnosus molybdenu
 m cofactor biosynthetic gene cluster, complete sequence. NID
 : g3955197.
 atgcaagaaagatactcgagacaagtgtgtttaaaggaaattgggtttaaaagggtcaaagt
 ctacttgagaaaaaacatgtgcttatagtaggtatgggtgcgttaggaacacacttagct
 50 gagggatttagtgagagcggaataaataagttaaccattgttgatagagattacattgaa
 ttagcaattttacaacgacaaacgctctttatagagcgtgatgcagaagatgtagctacca
 aaagtaatcgctgcacagaaagttttaaaagagatacgtaaagatgtagagatagatgct
 tatattgagcatgttaattacaatttcttagagcaacatggcatgcatgctgatatacata
 ttatagtgcaactgataattttgatacagctcagtttaattaatgactttgcttataaacat
 55 cagattccttggtttatgggtggtgtgtgtacaaagtacatatgttcaggcaacggtttatt
 cctggtgaaacaccgtgttttaattgcttaattgcctcaattaccatctattaatttaaca
 tgtgatacgggttgagttattcaaccagctgaacaatgacaaccagtttacaactcgtt
 gatgcattgaagtgtgctgactggttaataaggttaataaacacttcacttacggggatatt
 tggacaggagatcattatacatttggttttagtcgtatgcaaaatgaagattgtaaaact

tgtggtaatgctccaacatatccacaccttaatcaacatcaacaagattatgcgacctta
 tgtggaagagacactgttcaatataaaaatgctgatatttctcaggaaatattactatca
 tttctcgagcgaaatcatattcaatatcgacgaatttatatatgacaatgtttagggtt
 agagaacatcgaattgttgcattttctggaggtagatttttgatacatggaacgacagaa
 5 cctaaaaaagcaattcaattaatgcatcaactatttggttaa

Sequence 834

MQERYSRQVLFKEIGLKQSLLLEKKHVLIVGMGALGTHLAEGLVRAGINKLTIVDRDYIE
 FSNLQRQTLFIERDAEDVLPKVI AQAQVLKEIRKDVEIDAYIEHVNYNFLEQGHMVDII
 10 LDATDNFDTRQLINDFAYKHQIPWIYGGVVQSTYVQATFIPGETPCFNCLMPQLPSINLT
 CDTVGVIIQPAVMTTSLQLVDALKLLTGNKVNKHFTYGDWITGDHYTFGFSRMQNECKT
 CGNAPTYPHLNQHQDYATLCGRDTVQYKNADISQEILLSFLERNHIQYRTNLYMTMFRF
 REHRIVAFSGGRFLIHGTTEPKKAIQLMHQLFG*

15 Sequence 835

Contig_0520_pos_4068_5273,

is similar to (with p-value 0.0e+00)

>gp:gp|AF022796|AF022796_7 Staphylococcus carnosus molybdenu
 m cofactor biosynthetic gene cluster, complete sequence. NID
 20 : g3955197.

atgaaacaacatgttgaagtgaagaatatcaatattaatttagatgaaagtttaggacat
 attcttgctgaagatattgttgcgacctatgatataccaagatttaataaatcacctac
 gatgggtttgcaattagaagtgaagattcacaaggtgcaagtgccgaaaaccgtattgaa
 tttgaagtaatagatcatatcggtgcaggttcagtttcagaaaaaacaattgataaaaac
 25 caagcaattcgaataatgactggtgctcaaatcccttctggagctgatgccgtagtaatg
 tttgaacaaactattgaaatctgaaacaacttttacaattagaaaaatcctttaaacattta
 gaaaatatttctgctacaagggtgaagaaataaaagctggtgatattgtactacataaagg
 atgcgtattaactcaggtgtgatagcagctcttagctacatacgggtataactaaagtgcga
 gtggctcgaaaaccaactgttgcagtaattgctacaggtagtgaattgcttgaagtagaa
 30 gatgagcttgaaccaggaaagatacgaattcaaacggaccaatgattaaagcattagct
 aaacaatttggaaatacaagttggaatgtataaagttcagcatgataatctcgaaaagagt
 attgaggttgtaaaaaaagctttatcagagcatgatttagtaattactaccggaggtgtg
 tcggtaggagattttgattacttaccagaaatatacaagtcattccaagcacagataacta
 tttaaacaaagtggctcaaaagaccaggtagtgttactacggttgcatattgcagatggtaaa
 35 tatttatttggcttatctggaaacccttcagcctgctatacaggatttgaattatatgtc
 aaacctgctgtaaataagctcatgggagctaaagcttggtatccgcaaataatcaaagct
 acacttatggaagattttaataaagctaaccatttacacgattgattcgtgctaaggca
 acattaacaaaagctggaatgacagtaataccatctggatttaataaatcaggtgcagtt
 gtagccattgcgacagctaatagtatgattatgcttctcgtgggcacacgtggatttaaa
 40 gcgggcaacattgttgatgtgattttgaccgaatctaatagttttgaagaggaattgata
 ctatga

Sequence 836

MKQHVEVKININLDESIGHILAEDIVATYDIPRFNKSPYDGF AIRSEDSQGASGENRIE
 45 FEVIDHIGAGSVSEKTIDKNQAIRIMTGAQIPSGADAVVMFEQTIESETTFTIRKSFKHL
 ENISLQGEIEKAGDIVLHKGM RINSGVIAVLATYGYTKVRVARKPTVAVIATGSELLEVE
 DELEPGKIRNSNGPMIKALAKQFGIQVGMVKVQHDNLEKSIEVVKKALSEHDLVITGGV
 SVGDFDYLPEIYKSIQAQILFNKVAQRPGSVTTVAFADGKYLFGLSGNPSACYTGFEIYV
 KPAVNKL MGAKACYPQIIKATLMEDFNKANPFTRLIRAKATLT KAGMTVIPSGFNKSGAV
 50 VAIAHANAMIMLPGGTRGFKAGNIVDVILTESNSFEEELIL*

Sequence 837

Contig_0520_pos_5291_5746,

is similar to (with p-value 4.0e-40)

55 >gp:gp|AF022796|AF022796_8 Staphylococcus carnosus molybdenu
 m cofactor biosynthetic gene cluster, complete sequence. NID
 : g3955197.

atgaaaaattcagggaaaaccacattgatgaacatgctatatcatttttaaaagaacga
 ggctattcagtagtaacaattaacatcacgggcatatttggtgaagaaattgaattacag

tcattctgatgttgaccacatgaaacatttcgctgcgggcgagaccaaagtattgttcag
 gggcatcattttacagcaaacagtgcacgtaaaaaagaacaatcgcttagagaaataata
 gaaaattctgttacaattgattgtagtatcatttttagttgagggctttaagaagcaaat
 tatgataaaattatcgttttataaaaaataatgatgaattaagaagtctacaaggactttct
 5 caggtcatagggaaaatagaaaccaatcatccacgtgcaagtaatcaacttgagcactta
 ctcaataaaattaattaaggataaagggaatgaattaa

Sequence 838

MKNSGKTTLMNHAIISFLKERGYSVVTIKHHGHIGEEIELQSSDVDHMKHFAAGADQSIVQ
 10 GHHLQQTVTRKKKQSLREIIENSVTIDCSIILVEGFKEANYDKIIVYKNNDELRLSLQGLS
 HVIGKIEIETNHPRASNLQLEHLLNKLKDKGMN*

Sequence 839

Contig_0520_pos_5747_6199,
 15 is similar to (with p-value 6.0e-70)
 >gp:gp|AF022796|AF022796_9 Staphylococcus carnosus molybdenu
 m cofactor biosynthetic gene cluster, complete sequence. NID
 : g3955197.
 atgaagcaatttgaaatcgtagactcaacctattgaaacagaacaatatagggattttacg
 20 attaacgaacgtcaaggtgcccgtagtcgtatttactgggtcacgtaagagagtggaactaaa
 ggtatttcgtacacaacattttagagtagatgaagcttatataccaatggctgagaaaaaatta
 gctcaaattggtaagaaattgaagaaaagtgccctggaacaataacaacaattgtacat
 cgaattgggtccattacaaatatcagatattgcagttttaattgcagtatcttcaccgcat
 agaaaagcagcatatgcagcgaatgaatacgccatcgagcgcataaaaggaaattgttcca
 25 atttggaaaaaggaaatttgggaagatgggtgctgaatggcaaggtcatcaaaagggaaca
 tataatgaagcaaaaaagggaagcaagatga

Sequence 840

MKQFEIVTQPIETEYRDFITINERQGAUVVFTGHVREWTKGIRTOHLEYEAYIPMAEKKL
 30 AQIGKEIEEKWPGTITTIVHRIGPLQISDIAVLIAVSSPHRKAAYAANEYAIERIKEIVP
 IWKKEIWEDGAEWQGHQKGTYNKAKKGKAR*

Sequence 841

Contig_0520_pos_6555_7040,
 35 is similar to (with p-value 8.0e-39)
 >gp:gp|AF022796|AF022796_11 Staphylococcus carnosus molybden
 um cofactor biosynthetic gene cluster, complete sequence. NI
 D: g3955197.
 atgtttaatcgcatcattatcattagcactaattccaattagcttctcagtttgaatatgaa
 40 tatgtgattattgatgcgaacatcatcaaaataaagggccgctaacaggaatttactca
 gtgatgaacaatacatatgatgaagaattgtttttcattgtatctgttgatacaccaatg
 attacaagtaaaagcagtgaaatgggttatatcatttcattggtatcaaaacttaattgaatca
 cgtttagatattgtcgcattttaagaaggagaaatgtataccgacgattgggttttat
 acactttcgacgttttcttttattgaaaaagctttaaaattcaaatcatttaagtctgaag
 45 catgtctttaacaattatcgacagattgggttagatgttactgaaattgactcgccttat
 tattggtataagaatattaattttcagcatgatttggactctttaaaaatgcagataaat
 gaataa

Sequence 842

MFNRIIIISTNSQLASQFEYEVIIIDDEHHQNKGPLTGIYSVMKQYMDDEELFFIVSVDTPM
 50 ITSKAVNGLYHFMVSNLIESRLDIVAFKEGEICITIGFYTLSTFPFIEKALNSNHLSLK
 HVFKQLSTDWLDVTEIDSPYWYKNINFQHDLDLSLMQINE*

Sequence 843

Contig_0520_pos_7053_0,
 55 is similar to (with p-value 6.0e-63)
 >gp:gp|AF022796|AF022796_12 Staphylococcus carnosus molybden
 um cofactor biosynthetic gene cluster, complete sequence. NI
 D: g3955197.

- atgaaagaggtaataacaagataaattaggccgtccaatacgggatttaagaatatcggtc
actgatcgatgtaatttcagatgtgattattgtatgccaaaggaaatctttggagatgat
tacactttcttacctaagaatgaattgcttacttttgaagaattaacacgaatttcaaag
atztatgctcaattaggaggttaaaaagataagaattacaggaggagagcctctcttacga
5 cgcaatctttataaaactttagagcaattaaatctcatagatggtatagaggatattgga
ttgactactaatggcttgttattaaaaaacatggaaaaatttatatcaagctggttta
cgacgtattaatgtaagtttagatgcgattgaggataacgtttttcaagaaattaacaat
agaaatattaaagcgtctacaatcttagaacaattgattatgcagtatcaataggtttt
gaagttaaagtaaac
- 10 Sequence 844
MKEVIQDKLGRPIRDLRISVTDRCNFRCDYCMPKEIFGDDYTFLPKNELLTFEELTRISK
IYAQLGVKKIRITGGEP LLRRNLYKLVEQLNLIDGIEDIGLTTNGLLLKKHGKNLYQAGL
RRINVSLDAIEDNVFQEINNRRNIKASTILEQIDYAVSIGFEVKVN
- 15 Sequence 845
Contig_0520_pos_6172_5867,
is similar to (with p-value 5.0e-48)
>gp:gp|AF022796|AF022796_9 Staphylococcus carnosus molybdenu
20 m cofactor biosynthetic gene cluster, complete sequence. NID
: g3955197.
atgttccctttttagtgaccttgccattcagcaccatcttccaaatttcctttttccaa
ttggaacaatttcctttatgcgctcgatggcgtattcattcgctgcataatgctgcttttc
tatgcggtgaagatactgcaattaaaactgcaatatctgatatttgtaatggaccaattc
25 gatgtacaattgttgttattgttccagccacttttcttcaatttctttaccaatttgag
ctaattttttctcagccattggtatataagcttcataactctaaatggtgtgtacgaatac
cttttag
- 30 Sequence 846
MFPFDDLAIQHHLPKFPFSKLEQFPLCARWRIHSLHMLLFYAVKILQLKLQYLIFVMDQF
DVQLLLLFQATFLQFLYQFELIFSQPLVYKLHTLNVVYEYL*
- 35 Sequence 847
Contig_0521_pos_1712_2014,
is similar to (with p-value 1.0e-29)
>sp:sp|P37547|YABF_BACSU HYPOTHETICAL 20.7 KD PROTEIN IN MET
S-KSGA INTERGENIC REGION. >gp:gp|D26185|BAC180K_104 B. subti
lis DNA, 180 kilobase region of replication origin. NID: g46
7326. >gp:gp|Z99104|BSUB0001_41 Bacillus subtilis complete g
40 enome (section 1 of 21): from 1 to 213080. NID: g2632267.
atgtttgcatgttctattcctattttacctcttttacttttgcccttttctctatctaca
tatgctgtgttaacaccagaaacatgttcccgtatagtatcttcaattttatcacctggg
aaatctggatctgtcagtacaatcacacctctagtttgcgcgctgttgcatacattct
aaagtatatgtatcaattgcactaccgtttgtttcaatagtatcacaatctacagcactt
45 ttactcgttcagtatcatctttaccttcaacaacaataaattcgtttattttcataaat
taa
- 50 Sequence 848
MFACSIPIPLLLLLAFSLSTYACLTPECTSRIVFLILSPGKSGSVSTITPLVCCACCITS
KVYVSIALPFVSIVSQSTALFTRSVSSLPSTTINSFIFIN*
- 55 Sequence 849
Contig_0521_pos_8707_8264,
putative peptide of unknown function
atgtcaatagcaagaccggtagcagaaatgctaagtacctatacaactcaaataagaagca
gcatcgtcgttaaatgaggaatttgatttagttactttaagaaaatcaatcatcagatgg
tgtcaattaatattatctaataaagctatggctcttataggtgtcattgagttattaaaa
caagctaagaatcgtaattacaattactttactttatcagccgtcaatggctttttcgaa
gatataatgcatgcaaagatagaatggataattattatacatatttagtgatttaactgaa

gaaattgaaaattatgcaaatcaattaacttttaataatcaattaatcctcatgtatgatcag
attactgaagcgcataaaaagttaaatcaaaatgttaatccaacacttggttttgaacaa
atagtaataaaaaggtgtgatttaa

5 Sequence 850

MSIARPVAEMLSTYTTQIEAASSLNEEFDLVTLRKSIIRWCQLILSNKAMALIGVIELLK
QAKNRKLQLLTLSAVNGFFEDIMHAKIEMDNYTFSDLTEIENYANQLTFNQLILMYDQ
ITEAHKKLNQNVNPTLVFEQIVIKGVI*

10 Sequence 851

Contig_0521_pos_8262_7459,

is similar to (with p-value 3.0e-88)

>sp:sp|P37541|YAAT_BACSU HYPOTHETICAL 31.2 KD PROTEIN IN XPA
C-ABRB INTERGENIC REGION. >gp:gp|D26185|BAC180K_96 B. subtil
15 is DNA, 180 kilobase region of replication origin. NID: g467
326. >gp:gp|Z99104|BSUB0001_32 Bacillus subtilis complete ge
nome (section 1 of 21): from 1 to 213080. NID: g2632267.

atgccaatgtttaggtgttcagtttcaaaaagcagggaaattagaatactacgcgccg
aatcaattagatgtagaggttggtgactgggtgttgcctaatctaaaagaggtatagaa
20 attggccacgtaaaagttccattacgtgaagttgatgtagaagatgtcacattaccgcta
aaaaatatcatctcgtaaaatgaatgaagatgatcaagaaacatattatcgtaatgaacgc
gatgccaatgatgcgttagaattatgtaaaaagtagttaaagatcagcaattagatatg
cgattagtttaattgtgaatatacattagataaatctaaagtgatttttaattttaccgca
gatgatcgcatgtattttcgcaaacctgttaaagttttagctcaaaatctaaagactaga
25 atagaattacgtcaaatggggtaagagatgaagcgaaattattgggtggtatcggtcct
tgtggacgttctttatgttgttctacatttttaggagatttcgaacctgtatccattaaa
attggcgaaagatcagaacctatcattaaatccaactaagatttcaggagcttgtggtaga
ttgatgtgttctttaaataatgaaaatgactactatgaagaggctcgaactcaattacct
gatgttgagatatgattcaaacaccagatgggtcacggaaaagtgataggattaaatatt
30 ttagatattttctatgcaagttaaaatagaggtctagaacaacctttagaatataaaatg
gaagagatagaagtattgaattaa

Sequence 852

MPNVVGVOFQKAGKLEYAPNQLDVEVGWVWVQSKRGIEIGHVKFPLREVDVEDVTLP
35 KNIIRKMNEDDQETYYRNERDANDALELCKKVVKDQQLDMRLVNCEYTLDSKVIFNFTA
DDRIDFRKLKVLAKNLKTRIELRQIGVRDEAKLLGGIGPCGRSLCCSTFLGDFEPVSIK
MAKDQNLNPTKISGACGRMLCCLKYENDYEEARTQLPDVGDMIQTDPGHGKVIGLNI
LDISMQVKIEGLEQPLEYKMEIEVLN*

40 Sequence 853

Contig_0521_pos_7412_7095,

putative peptide of unknown function

atgaaattagaacatcacgttgaacaacttacaaccgacatgtcagaacttaaagattta
acaagtgaacttgttgaggagaatgttgctttgcaagttgaaaatgaaaatttaaacga
45 ttgatgaacaaaactgaagaatcggttgaaactcacttagataaagataattataagcat
gtaaaaacaccatctccaagtaagataatttagcaatgttatatcgtgaaggttttcat
atttgtaagggtgaattatttcgggaacatcgatcgtgaagattgcttattatgcctt
aatgtgttgagtgattaa

50 Sequence 854

MKLEHHVEQLTDMSELKDLTVELVEENVALQVENENLKRLMNKTEESVETHLDKDNKYH
VKTPSPSKDNLAMLYREGFHICKGELFGKHRHGEDCLLCLNLVLS*

Sequence 855

55 Contig_0521_pos_6999_6274,

is similar to (with p-value 2.0e-54)

>sp:sp|P37543|YABB_BACSU HYPOTHETICAL 28.3 KD PROTEIN IN XPA
C-ABRB INTERGENIC REGION. >gp:gp|D26185|BAC180K_98 B. subtil
is DNA, 180 kilobase region of replication origin. NID: g467

326. >gp:gp|Z99104|BSUB0001_34 Bacillus subtilis complete genome (section 1 of 21): from 1 to 213080. NID: g2632267.
 atgttagaagataatgaacgcgttgaccatttaataaaagaaggatatgagattatacaa
 aatgatgaagtattctcttttctactgatgccctattgttaggtatttaaccgaagt
 5 cgcaaaaatgataaagttatggatttatgctccggcaatgggtgttattccattattatta
 gctgctaaatcaactcaacctattgaaggaatagaaatacaagagcaactagtgtatg
 gcacgtcgagtttttaattgaatgatttgaacgatagactaactatgcacatatggat
 ttaaaagatgtatatcaaacatttcaacctgctcaatatacattagtgtacttgaatcct
 ccttatttttaaatgaatcaaatcatcaacatcaaaaagaagcacataaaatagcacgt
 10 cacgaaataatgtgtaatcttaaagattgtattgaagctgcaagacatttacttaaagag
 ggtggtcgttttattatggttcacgagcggaaaggctaattggatgtcttaaccgaatta
 agacatggtaaaattgagcctaaagcactgacgttagtgatagtaaacaatgataagcct
 gcacaaacaattgttgggaaggaagaaagggtggaaccaagggttagatatacgtaat
 ccattatacatatataatgaggatggatcatatagcgatgagatgaaagggtgttattat
 15 ggataa

Sequence 856
 MLEDNERVDHLIKEGYEIIQNDEVFSFSTDALLGLYTEVRKNDKVMDLCSGNGVPLLL
 AAKSTQPIEGIEIQELVSMARRSFKLNDLNDRLTMHHMDLKDVYQTFQPAQYTLVTCNP
 20 PYFKMNQNHQHKEAHKIHAEIMCNLKDCEAARHLLKEGGRFIMVHRAERLMDVLT
 RHGKIEPKALTLVYSKHKPAQTIVVEGRKGGNQGLDIRNPLYIYNEDGSYSDEMKG
 VYVG*

Sequence 857
 25 Contig_0521_pos_6031_5192,
 is similar to (with p-value 2.0e-67)
 >sp:sp|P37544|YABC_BACSU HYPOTHETICAL 33.0 KD PROTEIN IN XPA
 C-ABRB INTERGENIC REGION. >gp:gp|D26185|BAC180K_99 B. subtil
 is DNA, 180 kilobase region of replication origin. NID: g467
 30 326. >gp:gp|Z99104|BSUB0001_36 Bacillus subtilis complete ge
 nome (section 1 of 21): from 1 to 213080. NID: g2632267.
 atgacaactttatatttagtaggaacaccaattggaaatcttggtgatattacatttcga
 gctatagaacatttaaaaaaagttgatgtgattgcatgtgaagatacacgtgtaacacgg
 aaattgtgtaatcattatgaaatacaaacacctctaaagtcgtatcatgaacataataaa
 35 gaacaacaaactgactatttaatacagcagttacaaactggcttaaatatagcgtagta
 tcagatgctgggttgccattaattagtgatccaggatatgaattgggtgtcgaagcacgt
 aaaaataatataaatatagaacagtagcaggtcctaattgctgggttgactgcacttatg
 tcaagtggtattaccatctttcacatacacatttttaggtttttgccaagaaaagaaaaa
 gaaaaaattgaagtgttgaggatagaatgtttcaaaatagtactttaataactttatgaa
 40 tcgccttatagggttactgatactttgaaagcaatagctaaaaatagattcacaaagatgg
 attactgttggttagagagctaacgaagaaatttgaacaagttcttacacttacagttgat
 gatattgtgaaattgattaatcatgacaaattacctcttaaaaggtaggtttgtgatactg
 attgaagggtgcattacctaagagtggtgaatcatgggtttgaaagctatacggttaaagaa
 catgttgattattatattgaaaccaaacatgttaaacctaaaaaagcaattaaatttgct
 45 gctacagatcgacatatgaagacgggtgacatatataatatttatcataatattgattaa

Sequence 858
 50 MTTLYLVGTPIGNLGDITFRAIETLKKVDVIACEDTRVTRKLCNHYEIQTPPKSYHEHNK
 EQQTDYLIKQLQTGLNIALVSDAGLPLISDPGYELVVEARKNNINIETVPGPNAGLTALM
 SSGLPSTYTFGLFLPRKEKEKIEVLEDRMFQNSTLILYESPYRVDTLKAIAKIDSQRW
 ITVGRELTKKFEQVLTTLTVDDMLKLINHDKPLKGEFVILIEGALPKSGESWFESYTVKE
 HVDYYIETKHVKPKKAIKFVATDRHMKTGDIYNIYHNID*

55 Sequence 859
 Contig_0521_pos_4868_3012,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P23920|SYM_BACST METHIONYL-TRNA SYNTHETASE (EC 6.1.1.
 10) (METHIONINE--TRNA LIGASE) (METRS). >pir:pir|S16682|S1668

2 methionine--tRNA ligase (EC 6.1.1.10) - *Bacillus stearothermophilus* >gp:gp|X57925|BSMETSG_1 *B. stearothermophilus* metS gene for methionyl-tRNA synthetase. NID: g39988.

5 atgcaaggctatgatgttcgttatttaactggcactgatgagcacggtcaaaaaatccaa
gaaaaagctcaaaaaagctggcaaacagaactagaatacttagatgaaatgatttcagg
attaaaaacttatggagtaaaacttgagatttctaataatgatgattttattcgaactacagaa
gagcgtcataagcaagtcgttgagaaagtgttgagcgattattaaaacaaggtgacatt
tatttaggtgaatacgaaggttggtattctgttcctgatgaaacatattatacagagtca
caacttgttgaccctgtttatgaaaacggcaaaattgtagggtgtaaaagtccctgattct
10 ggtcacgaagtcgaacttgtaaaagaagaaagctatttcttcaacattaataaatatataca
gaccgcttattagaattttacgatgaaaatccagactttatacaaccaccatctagaaaa
aatgaaatgatttaactttatcaaacagggtttagaagatttagcagtatcacgtaca
tcattcgcattgggtgtacgtgtaccatctaatacctaatacatgttgatatacgtgtgatt
gatgcacttggttaattatatttcttcattaggttatctatctgatgatgaaacattattt
15 aataaatattggccagcagacatacacttgatggctaaagaaattgtacgtttccactct
attatatggccaatattgttaatggcgttggtattaccacttcctaaaaaagtttttgca
cacggttggtatttaataagagatggtaaaatgagtaaatctaaaggaatgtcgtatgat
cctaatagtattaattgatcggttatggtcttgatgacgacagcttattacttaatgcgtgaa
ttaccgtttggtttctgatggcgattttacacgggaagcctttggtgaaagaacaaattac
20 gatcttgcgaaatgatttaggtaattcagtgaaatcgtaactatctctatgataaacaatat
ttccacggcgaattacctgcataccaaggtccaaaacatgaattggatgaaaaaatggaa
gcgatggcgcttgaaactgttaaatcattcaatgataatatggaaagtttacaattttct
gttgctttatcaacagtatggaaattattagtcgtacaaacaatatattgatgaaact
caaccttgggttcttgcaaaagatgaaaatcaacgtgagatgcttggtaatgtaattggca
25 catcttgctcgagaacattcggtttcgtacaaatcttattacaaccattcttgacgcatgca
cctagagaagattttaagcaacttaattattaacaatccggatttacatcaattagatagt
ctgcaacaatatggtatgttgctcagaggcaattactgtaactgaaaagccaacaccaatt
ttcccaagatttagacactgaagcagaaattgcttatatcaaagaatcaatgcaaccacct
aaatcaataaaaacagtcctgatgaacccggtaaagagcaaatgatatacaagattttgat
30 aaagttgaaatcaaagcagcaaccattattgatgcggaataatgtaaaaaatcggagaaa
ctattaaaaataaaagttgaattagataatgaacaacgtcaaatagtatctggtatagct
aagttttatcgctccggaagacattattggtaaaaaagttgcagttgttactaatttaaaa
ccagctaaattgatgggacaaaatccgaaggtatgattttgtcagctgaaaagatggc
gtacttaccttgataagcttgccctagcgcaattccaaatgggtgcagtaattaaatag

35 Sequence 860
MQGYDVRYLGTGDEHGQKIQEKAQKAGKTELEYLDEMISGIKNLWSKLEISNDDFIRTE
ERHKQVVEKVERLLKQGDIIYLGEYEGWYVDPDETYTESQLVDPVYENGKIVGGKSPDS
GHEVLVKEESYFFNINKYTDRLLEFYDENPDFIQPPSRKNEMINNFIKPGLEDLAVSR
40 SFDWGVVRVPSNPKHVVVWIDALVNISSLGYLSDDETLFNKYWPADIHLMAKEIVRFHS
IIWPILLMALDPLPKKVFAHGWIWMKDGKMSKSGNVVDPNVLIDRYGLDATRYLLMRE
LPFGSDGVFTPEAFVERTNYDLANDLGNLVNRTISMINKYFHGELPAYQGPKHELDEKME
AMALETVKSFNDNMESLQFSVALSTVWKFI SRTNKYIDETQPWVLAKDENQREMLGNVMA
HLVENIRFATILLQPFLLTHAPREIFKQLNINNPDLHQLDLSLQYGMMLSEAITVTEKPTPI
45 FPRLDTEAEIAYIKESMQPPKSIKQSDPEFGKEQIDIKDFDKVEIKAATIIDAENVKKSEK
LLKIKVELDNEQRQIVSGIAKFYRPEDIIGKKVAVVTNLKPAKLMGQKSEGMILSAEKDG
VLTILSLPSAIPNGAVIK*

Sequence 861
50 Contig_0521_pos_2987_2214,
is similar to (with p-value 3.0e-99)
>sp:sp|P37545|YABD_BACSU HYPOTHETICAL 29.2 KD PROTEIN IN MET
S-KSGA INTERGENIC REGION. >gp:gp|D26185|BAC180K_102 *B. subtilis* DNA, 180 kilobase region of replication origin. NID: g46
7326. >gp:gp|Z99104|BSUB0001_39 *Bacillus subtilis* complete g
55 enome (section 1 of 21): from 1 to 213080. NID: g2632267.
atgatgttaatcgatacgcacgtacatttaaatgatgaacaatatgatgaggatttaaat
gaagtgatattctcgtgcgagagaagcaggcgtagatagaatgtttgtagtaggtttgat
acacctactattgaacgtactatggagctaataagataagtatgactttatctatggtatt

atcggttggcatcctgttgatgcaatagattgtactgatgaaagattggaatggatagaa
 agtcttttctaaacatcctaaaaattattgggtattgggtgagatggggttagattatcattgg
 gataaatcaccttctgatgtacaaaaagaggtatttaaaaagcaaattgcattagctaaa
 cgtgttcaattacctattattattcataatcgtgaagcgactcaagattgcatagatatt
 5 ttgattgaagaacatgcagaagaagtgggcggaataatgcatagtttttagtgcttcacct
 gaaattgctgatgtcgtgattaataaattgaacttctatgtttcgcttgaggaccgctc
 actttcaaaaatgcaaaaacaacaaaagaagtgtctaaacacgtaccaatggatcgtttg
 ttatgcgagacagatgccccgtatctatccccgcacccttatagaggtaaacgtaatgaa
 ccagaacgtgttacttttagtagcacaacaaattgcagatttgcgtgggtatgacttatgaa
 10 gaggtctgtcgccaaacaaccgaaaatgctgaacgtttattcaatttgaattaa

Sequence 862

MMIDITHVHLNDEQYDEDLNEVISRAREAGVDRMFVVGFDPTTIERTMELIDKYDFIYGI
 IGWHPVDAIDCTDERLEWIESLSKHPKIIIGIGEMGLDYHWDKSPSDVQKEVFKKQIALAK
 15 RVQLPIIIHNREATQDCIDILIEEHAEEVGGIMHSFSASPEIADVINKLNFYVSLGGPV
 TFKNAKQPKVAKHVPMDRLLCETDAPYLSHPYRGRKNEPERVTLVAQQIADLRGMTYE
 EVCRQTTEAERLFLNLN*

Sequence 863

20 Contig_0521_pos_2007_1462,
 is similar to (with p-value 2.0e-43)
 >sp:sp|P37547|YABF_BACSU HYPOTHETICAL 20.7 KD PROTEIN IN MET
 S-KSGA INTERGENIC REGION. >gp:gp|D26185|BAC180K_104 B. subti
 lis DNA, 180 kilobase region of replication origin. NID: g46
 25 7326. >gp:gp|Z99104|BSUB0001_41 Bacillus subtilis complete g
 enome (section 1 of 21): from 1 to 213080. NID: g2632267.
 atgaaaataaacgaatttattgttgttgaaagtgatgactgaacgagtaaaaagt
 gctgtagattgtgatactattgaaacaaacgtagtgcaattgatacatatacttttagaa
 gtgatacaaacacgcgcagcaaaactagaggtgtgattgtactgacagatccagatttccca
 30 ggtgataaaaattagaaatactatacgggaacatgtttctggtgttaaacacgcatatgta
 gatagagaaaaaggccaaaagtaaaagaggtaaaataggaatagaacatgcaaacattaaa
 gatattcaagaagcattaatgcatgtaagttcaccacttgaagaagctaaagaaactatt
 gataaaaagtgtactcattgttgggatttaattatcggtaaagatgcaagataccgtaga
 aatatcttaggtcgaaaattacacatcggtcactctaattgaaagcaattattaaagaaa
 35 cttaatgcttttggctatactgaagacgatgtcagaaaagcgctatttgaagaagaggag
 aattaa

Sequence 864

40 MKINEFIVVEGKDDTERVKSADVCDTIETNGSAIDTYTLEVIQHAQQTRGVIVLTDPDFP
 GDKIRNTIREHVSQVGHAYVDREKAKSKRGKIGIEHANIKDIEALMHVSSPLEEAKETI
 DKSVDLIDGLIIGKDARYRRNILGRKLHIGHSNGKQLLKKLNAFGYTEDDVRKALFEEEE
 N*

Sequence 865

45 Contig_0521_pos_1461_571,
 is similar to (with p-value 2.0e-99)
 >sp:sp|P37468|KSGA_BACSU DIMETHYLADENOSINE TRANSFERASE (EC 2
 .1.1.-) (S-ADENOSYLMETHIONINE-6-N', N'-ADENOSYL(RRNA) DIMETH
 YLTRANSFERASE) (16S RRNA DIMETHYLASE) (HIGH LEVEL KASUGAMYCI
 50 N RESISTANCE PROTEIN KSGA) (KASUGAMYCIN DIMETHYLTRANSFERASE)
 . >gp:gp|D26185|BAC180K_105 B. subtilis DNA, 180 kilobase re
 gion of replication origin. NID: g467326. >gp:gp|Z99104|BSUB
 0001_42 Bacillus subtilis complete genome (section 1 of 21):
 from 1 to 213080. NID: g2632267.
 55 atggaatataaagatatagcaacaccatctcgaacacgtgctttgcttgatcaatatggg
 ttttaattttaagaaaagtttaggacaaaattttctaatagatgtaaatatcattaataaaa
 attatcgaagcgagtcataatagattgtacaacgggtgtaattgaagttggaccaggtatg
 ggatcattgactgaacaacttgcaagaatgctaagaaggtgatggcttttgaaattgat
 caaagattaatacctgtgcttaaaagatacactttcaccatacgataatgtaacaattatc

aatgaagatatactttaaagctgatattgctaagctgtagatacacatctacaagattgt
 gacaagattatgggtgttgctaatttaccgtattatattaccacacctattttacttaat
 ttgatgcaacaggatgtacctattgatggttttgtcgtaatgatgcaaaaagaggttagga
 5 gaacgtttgaacgctcaagtaggtaccaaagcatagcgttcgttatcgattgttgctcaa
 tactatacggagacaagtaaaagttttaacagttcctaaaactgtatttatgcctcctcca
 aacgttgattctatcggttgtaaaattgatgcaacgccaagaaccacttgtagaggttgat
 gatgaggaaggcgttttttaagtttagcaaaaggccgcttttgcacaacgacgtaaaacaatt
 aataataactacccaaaacttctttaaagatggtaagaagaataaagaaactatacgacag
 10 tggctagaaagcgtggtattgatcctaaaagacgtggagaacactcacgattcaagat
 ttcgccacattatatagaacaaaagaaaaattctccgaattaacaaattaa

Sequence 866

MEYKDIATPSRTRALLDQYGFNFKKS LGQNFLIDVNI INKII EASHIDC TTGVIEVGPGM
 GSLTEQLAKNAKKVMAFEIDQRLIPVLKDTLSPYDNVTI INEDILKADIAKAVDTHLQDC
 15 DKIMVVANLPYYITTPILLNLMQQDVPIDGFVMMQKEVGERLNAQVGT KAYGSL SIVAQ
 YYTETSKVLTVPKTVFMPPPNVDSIVVKLMQRQEPLVQV DDEEGFFKLAKAFAQRRKTI
 NNNYQNFFKDGKKNKETIRQWLESAGIDPKRRGETLTIQDFATLYEQKKKFSELTN*

Sequence 867

20 Contig_0522_pos_1721_315,
 putative peptide of unknown function
 atggcaattcctaaccattatcaactacggattggcatgcatataaagttaatctaagt
 caatatttgactcaagaaaatggcgttatttaggacatttattgaatgggttgccgta
 cataataacataataaagagctttaatatatgcgataaacttcgttttttagttatctattta
 25 gttgcttatatggttcaattacatacgaatcgtattttattttatttttagttttgtgta
 atggttactgtacctataacaattttatagcgaaacttacgggtgggttactggatttttt
 agttatatacctgctacagtcctatcactttttattctttttacggtagttaaaaagatt
 gagtcgcacgatacagtttctgaaatgcaattatgggtattttttatttagtaagttgttt
 ggacaattcttcttgagaatctttccatcgctaataagcttaatttttaataaggaatg
 30 gtagtctattttcttggtaaaaaaagactcagttatttcttaattgtaggatttatgctt
 agttgtataggttaacattataatgtttttaaacttcaattattttttaattaaggatgga
 ttaaaatacgcattattcaatttccgatagtcaggaatgatacataaagcaggtgtgacg
 ttatttaagctttaccagaatatgtttattaatcaaatgattattcttaccgtgata
 tcaatagtaagtatagttttacttaagcaaaaataaaagcctgaagcatatgagagtttat
 35 attaaaaataccactactcttaggtttaattactttacattttataagatcttcgtttac
 aatcaatttcatatttgaattatataaagcttcattttctatagccgttttgaatacaacg
 atttgcttcatttatacatgataagtgatatacgttggtgtttaaaatgatacagcaaaga
 tacataagaatgattgtgatggggagttttatagctatggcttcactgttttgccactt
 ttatttgtgacgcctataagttatagaattttttattttattttatactttatggatcgtg
 40 ttatttactttgttaattcagcaatgtgatgtgctatttaaaacaacttgaacatataatt
 aaaatatttgcgattatcatcagcatcattatgatgattggatttacttttatacatatt
 agtagtggtcacagaatagacttcattaaagaacaaataacacaacatcatcgctatcag
 aaaataacatttgaaagattaccatttgagcgatatactcatatgactacaccaaagtcg
 aaggaacaacttcaagatttcaaacactattatgatttgcccaaagacatcacatttaaa
 45 gtagtcccatatggtacaaaacaataa

Sequence 868

MAILTPLSTTDWHAYKVNLSQYL TQENGRYLGH LFEWVAVHNNIIRALIYAIT SFLVIYL
 VAYMVQLHTNRIYFILSFVLMVTPNTIYSETYGWFTGFFSYIPATVLSL FILFTVVKKI
 50 ESHDTVSEMQLWVFLVSLFGQFFLENLSIANS LIILIGMVVYFFVKKRLSYFLIVGFML
 SCIGNIIMFLNFNYFLIKDGLNTHYSISDSHGM IHKAGVTLFKLVPEYMFINQMIILTVI
 SIVSIVLLKQNKSLKHM RVYIKIPLLLGLITLPIYKIFVYNQFHFELYKASF SIAVLNTT
 ICFIYMISVIYVVF KMIQQR YIRMIVMG SFIAMASSVLP LLEVTPISYRN FYFIYTLWIV
 ILLCLIQQCDVLFKQLEHI IKIFAI IISIIMMIGFTFIH ISSVHRIDFIKEQITQHHR YQ
 55 KITLERLPFERYTHMTTPKSKEQLQDFKHYYDL PKDITFKVVPYGTQK*

Sequence 869

Contig_0523_pos_5280_0,
 is similar to (with p-value 4.0e-55)

>pir:pir|B26532|B26532 tyrA protein - *Bacillus subtilis* >gp:
 gp|M80245|BACVARGNS_17 *B. subtilis* dbpA, mtr(A,B), gerC(1-3),
 ndk, cheR, aro(B,E,F,H), trp(A-F), hisH, and tyrA genes, co
 mplete cds. NID: g143798. >gp:gp|Z99115|BSUB0012_201 *Bacillu*
 5 s *subtilis* complete genome (section 12 of 21): from 2195541
 to 2409220. NID: g2634478.
 atgactcatatacaatattgttattctgaaaattctgaaacaaggaggcgctctatgcga
 aatattttattttagtgagggttaggccttattggcggtagcttggcgagtaattttaaataat
 cattacagtaatttcaatattcttgcatacattcggactacacacaacttgatgaagcc
 10 ctttctataggtattattgatcaaaaaggttaattgattatgctactgctggtgagatagcg
 gatataatcatctttgcaactcctgttgagcaacaattaaatatctatctgaacttaca
 aattacaatacaaaaactcatttgattgtaacagacacaggtagtagtaccaaaacttactata
 caatcattcgaaaaagaattattaaaacatgatattcatttaattagtggtcatcctatg
 gcaggaagtcataaatctgtgtgttttaaacgcgaaaaaacatttatttgaaaatgcttat
 15 tacattctgtatttgaatgaaatcgaaaaataatgaagccgcgacatattttaaagaaatta
 cttaaacctacgttagcaaaaatttatcgttactcatgcaaatgaacatgatttcgtaacc
 ggtatagtgagtcattgttccacatatcatcgcttcaatttttagttcatctaagtgcataat
 catgtcaaaagaccattctttaatcgaaaaattagcagccggtggcttttagagatataact
 cgtatagcaagtagtaattgctcagatgtggaaggatatcacttttaataatcaaaatcat
 20 attttatctttacttaacgagattaaagaacaaattactggtattgaaaatttgata

Sequence 870

MTHIQCYSENSETRRRSMRNILFVGLGLIGGSLASNLKYHYSNFNILAYDSYTLQLEA
 LSIGIIDQKVNDYATAVEIADIIIFATPVEQTIKYLSELNTYNTKTHLIVTDTGSTKLT
 25 QSFKEKLLKHDHILISGHPMAGSHKSGVLNAKKHLFENAYYILVFNEIENNEAATYLLKL
 LKPTLAKFIVTHANEHDFVTGIVSHVPHIIASILVHLSANHVKDHSLIEKLAGGFRDIT
 RIASSNAQMWKDITLNNQHILSLLNEIKEQITGIENLI

Sequence 871

30 Contig_0523_pos_4745_3273,
 is similar to (with p-value 2.0e-78)
 >sp:sp|Q02001|TRPE_LACLA ANTHRANILATE SYNTHASE COMPONENT I (
 EC 4.1.3.27). >pir:pir|S35124|S35124 anthranilate synthase (
 EC 4.1.3.27) alpha chain - *Lactococcus lactis* subsp. *lactis*
 35 >gp:gp|M87483|LACTRPOP_2 *L. lactis* trpE, trpG, trpD, trpF, t
 rpC, trpB trpA genes, complete cds. NID: g149514.
 gtgggtaccgcgcgtcagcgtccttataacgaaggaggctgatttttttcggaaaaggagg
 aatcaattggatattgtatcaaaaaggtgaatgctcaaattacgccagaagccttagca
 aaattaaaaacaaaaaagatcatttttgaaagtacaaatcaacagaaaacttaaaggtagg
 40 tactcgatagtagtattcgatcattatggcaaaattacattagataattctcaactttta
 attaagttagacaattcattgtgaaatagtttaagaatcaaccgtatcaacgacttaaggaa
 tttgtagataaatattattttgaaatcaaaagataaatattttaaagatttaccttttatt
 tcgggctttatagggacatgtagctttgatttagtagcatgaatttaaaaaattacaa
 gatattaaattagaagatcatcaaaactcatgatgtccaattttatctagtggaaagatgta
 45 tttgtttttgatcattataaagatgaattatatattatcgcaagtaacttattttcttat
 agaacaaaagagagattaaaggaatctattgaacgtaaaattgaagatttaaaaaacata
 catttttcggttgaggatataaattataaatccatccctcgacatataaccaccaatata
 tcagagcaacaattttgttcaaaactattagaatttttaaaaaagaaaattactgaaggagat
 atgtttcaagtagttccttcaagaattttatagttataaacaccattttcaacacaattta
 50 catcaattaaacttttcagttatatcaaaaatttaaagcgacaaaatcctagtccatatatg
 tattatattaataaagatgtaccgattgtaataggaagttctcctgaaagttttgtaaag
 gtaaaagatggaaaagtttatacgaatcctatagctggaacaattaaaagaggtcaaaat
 aaaaaagaagatgaaaataatgaaaagacattaatgaaagatgaaaaggaattgagtga
 catcgtatgctcgtagatttaggaagaaatgatattcatcgaataagtaaaacaggcact
 55 tcacaaattaccaaactaatgacaatagaacgttatgaacatgcatgcatatcgttagt
 gaagttattggagaattaaaaccccatctatctcctatgagcgtcatcgcaagtttgcta
 ccaacgggtactgtctcaggtgcacctaaacttagagctatacagagaatatacgaatct
 tacccttataaaagaggtatctatagcgggtggtgttggttatatcaactgtaatcatcat
 ttagattttgcattggctatagctaccatgattatcgatgaggaaaaagtcagtgctcgag

gcaggatgtggagtagtatatgattctattccagagaaagaacttgaagaaacaaaactt
aaaagctaaaagtttattggaggtaactccatga

Sequence 872

5 VVPRVSVLITKEADFFRKRKRSMDIVYKKVNAQITPEALAKLKQKKIIFESTNQQLKGR
YSIVVFDHYGKITLDNSQLLIKLDNHCEIVKNQPYQRLKEFVDKYYFEIKDKYLDLPFI
SGFIGTCSFDLVRHEFKKLQDIKLEDHQTHDVQFYLVEDVFDHYKDELYIIASNLFSY
RTKERLKESIERKIEDLKNHFSVEDINYKSIPRHITTNISEQQFVQTIRILKKKITEGD
10 MFQVVPSRIYSYKHHFQHNLHQLTFQLYQNLKRQNPSPYMYINKDVPVIGSSPESFVK
VKDGVYTNPIAGTIKRGQNKKEDEENNEKTLMKDEKELSEHRMLVDLGRNDIHRISKGT
SQITKLMTIERYEHVMHIVSEVIGELKPHLSPMSVIASLLPTGTVSGAPKLRAIQRIYES
YPYKRGYISGGVGYINCNNHLDFAAIRTMIIDEEKVSVEAGCGVYDSIPEKELEETKL
KAKSLLEVTP*

15 Sequence 873

Contig_0523_pos_2707_1712,
is similar to (with p-value 8.0e-51)
>sp:sp|P17170|TRPD_LACCA ANTHRANILATE PHOSPHORIBOSYLTRANSFER
ASE (EC 2.4.2.18). >pir:pir|S42343|JS0340 anthranilate phosph
20 horibosyltransferase (EC 2.4.2.18) - Lactobacillus casei >gp
:gp|D00496|LBATRP_2 Lactobacillus casei DNA, trp operon (trp
D, trpC, trpF, trpB, trpA), complete cds. NID: g216754.
atgacccttcttgagaaaattaaacaaaataaatcttatctaaaaaagatatgcaatca
tttattgttacactgtttgattcaaatatagaaaccaatgtaaaggtgaattattgaaa
25 gcttatacaaataaagacatgggtcaatatgagctaactgatttagttgaatattttatc
cagacaaactatccaaaccaaccattttataataaagctatgtgtgtttgtggcacaggt
ggagatcaatcaaatagctttaatttctacaactgtagcttttgttagcaagtgc
ggagtgcagtcattaaacacggtaataaaaagtattacttcacattcaggaagtacagat
gtattacatgaaatgaatataaaaaacaaacaaatgaacgaagtagagcaacaattaaat
30 ttgaaaggattagcattcataagtgcactgattcttatccaatgatgaaaaagcttcaa
tcaattagaaaatcgattgcaacacctacaatttttaacttgattggaccattaattaat
cctttcaaattaacttatcaagtgtgggggtatatgaagcttcacaacttgaaaatata
gcacaaacattaaaaggatttaggtagaaaacgagcaattttaattcatgttgcaaatggg
atggatgaggccacgcttctgtggtgaaaatatcatttatgaagttagcagcgaaagagca
35 ttaaaaaaatatagtttaaaagcagaagaagtcggttttagcttatgcaataatgacacg
ttgataggtggttcacctcaacaaataaacaattgcattgaatatcctaagtggcacg
gatcactcaagtaaacgagatgtagtttgttaaatgctggaattgctttatatgttgct
gagcaagtggaaagtatcaaacatggcgtagagagagcgaaatatctcattgatacaggt
atggcaatgaaacaatatttaaaaaatgggaggttaa
40

Sequence 874

MTLLEKIKQNKSLSKKDMQSFIVTLFDSNIETNVKVELLKAYTNKDMGQYELTYLVEYFI
QTNYPNQPFYKAMCVCGTGGDQSNFSNISTTVAFVVASAGVPVIKHGKNSITSHSGSTD
VLHEMNIKTNMNEVEQQLNLKGLAFISATDSYPMMKKLQSIRKSIATPTIFNLIGPLIN
45 PFKLTYQVMGVYEASQLENIAQTLKDLGRKRAILIHGANGMDEATLSGENIIEVSSERA
LKKYSLKAEVGLAYANNDTLIGGSPQTNKQIALNILSGTDHSSKRDVLLNAGIALYVA
EQVESIKHGVRAKYLIDTGMAMKQYLMGG*

Sequence 875

50 Contig_0523_pos_1708_920,
is similar to (with p-value 7.0e-43)
>sp:sp|Q01999|TRPC_LACLA INDOLE-3-GLYCEROL PHOSPHATE SYNTHAS
E (EC 4.1.1.48) (IGPS). >pir:pir|S35127|S35127 indole-3-glyc
erol-phosphate synthase (EC 4.1.1.48) - Lactococcus lactis s
55 ubsp. lactis >gp:gp|M87483|LACTRPOP_5 L. lactis trpE, trpG,
trpD, trpF, trpC, trpB trpA genes, complete cds. NID: g14951
4.
atgactatttttaaatgaaattattgagtataaaaaaacttttgcttgagcgtaaatactat
gataaaaaacttgaaattttacaagataacggaaatgttaagaggagaaagctgattgat

tcacttaactatgatagaacattatcagttattgctgaaataaaatcgaaaagcccatct
 gtacctcaattaccgcaacgtgatcttggtcaacaagttaaagattatcaaaaatattggt
 gctaattgctatttcaatattaactgatgaaaaatactttggcggtagttttgaacgatta
 aatcagttatcaaagataacatcgttaccagttttatgtaaagattttattattgataaa
 5 attcaaatagatggttgcaaacgagctggtgcatctatttttattaatagtaaatatt
 ttaagtgatgaccaattaaaagaattgtattcatatgcaacaaaccataatttagaagct
 ctagtagaagttcatacaattagagaacttgaacgtgcacaccaaattaaccctaaaatt
 attggtgttaataatcgtgatttaaaacgatttgaaaccgatgttctacatacaataaaa
 ttacttaagtttaaaaagcttaattgctgctacatttcagagagtggtcattcatacaaaa
 10 gaagatggttgagaaaatagtagattcaagattgacggtttactttagtagggaggcatta
 atgaaaacaaatgacttaagtcagtttttgcttagtttaaaagttaaagaagaatctctat
 gatagttaa

Sequence 876

15 MTILNEIIEYKKTLLERKYYDKKLEILQDNGNVKRRKLIDSLNYDRTL SVIAEIKSKSPS
 VPQLPQRDLVQQVKDYQKYGAN AISILTDEKYFGGSFERLNQLSKITSLPVLCKDFI IDK
 IQIDVAKRAGASII LLIVN ILSDDQLKELYSYATNHNLEALVEVHTIRELERAHQINPKI
 IGVNNRDLKRFETDVLHTNKLKFKKSNCYI IESGIHTKEDVEKIVDSSIDGLLVGEAL
 MKTNDLSQFLPSLKLKKNLYDS*

Sequence 877

Contig_0523_pos_786_307,
 putative peptide of unknown function
 25 gtgccagatcatatagagaaaagtagtggtcgtagtaaatcctcaaattgtccaccataaag
 agaataattaatcaaactgatattaacacaatccaattacatggaaatgaaagcattcaa
 ttaattagaaatattaagaaacttaattcaaaaaataagaatcataaaagcaattccagca
 acaagaaatttaataataaacattcaaaagtataaaagatgagatagacatgtttattata
 gatacaccatcaatcacatacggaggagacaggtcaaagtttgactggaaattattaaaa
 aaaataaaggcggttgattttctcattgcggttggtttggattttgaaaagataaaacga
 30 ttagaaatatattcatttgacaatgtggttatgacatctcaactggcattgagtcacat
 aatgaaaaagattttaataagatgactcgaatattaaaatttttgaaaggagacgaatga

Sequence 878

35 VPDHIEKVVVVVNPQMSTIKRIINQTDINTIQLHGNESIQ LIRNIKKLNSKIRIIKAIPA
 TRNLNNNIQKYKDEIDMFIIDTPSITYGGTGQSFDWKLKKIKGVDFLIAGGLDFEKIKR
 LEIYSFGQCYDISTGIESHNEKDFNKMTRILKFLKGDE*

Sequence 879

40 Contig_0524_pos_471_1280,
 is similar to (with p-value 7.0e-48)
 >gp:gp|L19300|STAORFPHI_2 Staphylococcus aureus DNA sequence
 encoding three ORFs, complete cds; prophage phi-11 sequence
 homology, 5' flank. NID: g310601.
 45 atgaaaagaagcgataaatatacggatgattatattgaacaacgttatgagtcctcaacga
 ccttattacaatacatattatcaaccaatagggaaccacgaaaaagaaaaaagtaaa
 agaattttcttaaaagcaattatcactatattaattttattgattatatttttgggtgct
 atgtactttatttcttcaagagcaaatgtagatgattttaaatacaattgaaaataaaagc
 gattttgttgctaccgaaaatatgcctaactatgtaaaaggcgcatatttttcaatggag
 50 gatgagcgtttctataaacatcatggctttgatataaaaggaacgacaagggcattgttt
 tcaactattagcgatagagatgtgcaaggtggaagtacaattacgcaacaagttgtaaag
 aattattattacgataatgaacgatcctttacaagaaaaatcaaagaattgtttgtagcg
 cgtaaagttgaaaagcaatacagtaaaaatcagattttaagtttctatatgaataatatt
 tattattggtgataatcaatatactgtagaaggtgctgcaaatcattattttggtgtaacg
 55 ttcgataaaaaacaattcaaatatgagtcagatttagtggttcaaaagtgctatattagca
 agcaaaagtaaatgcaccaagtggtatgatgtaaatgatatgtcgaataattacatcaat
 agagttaaaaccaatttagagaaaatgaaacaacaaaattttatttagtgaatcacatatt
 caagaagctatgtctcaacttggaattaa

Sequence 880

MKRSDKYTDDYIEQRYESQRPYYNTYYQPIGKPPKKKKSKRIFLKAIITILILLIIFFGV
 MYFISSRANVDDLKSIENKSDFVATENMPNYVKGAFISMEDERFYKHHGFDIKGTTRALF
 STISDRDVQGGSTITQQVVKNNYYDNERSFTRKIKELFVARKVEKQYSKNQILSFYMNNI
 5 YYGDNQYTVEGAANHIFGVTVDKNNNSMSQISVLQSAILASKVNAPSVYDVNDMSNNYIN
 RVKTNLEKMKQQNFISESQYQEAMSQLGN*

Sequence 881

Contig_0524_pos_1558_2361,
 10 putative peptide of unknown function
 atgccaaaggttaactaaaatagaagtacaaaaaagaataaagaacgctttaatctcttt
 ttagatggagaatttgaaatgggatagatattgatacattagttaaatctaacttaaaa
 aaagatcaaatacttgaaacgtcagatatgcagaatattcaagaatatgatcactaccgt
 cgaggtgttaattctgcaattcaatacttgtcttataagaaacgtactgaaagagaagtt
 15 atacagttatttagaaaaaacgatattcaaagtaattgctattcaagatgtcattgactat
 tgctataaggaaaaatttattgatcatgaagactacgcagaaagtttaaaaaacaccatg
 atacacactacagataaaggaccagaaatatatagacaaaaactctatcaattaggtatt
 gaagttacgattattgaaaaatatgtcgaagcatatgaacaacaacaaccattagatgac
 gtcataaaagttgtgtaaaaagtgtatgaagcttaaaaagggtcctgaagcaaaaggtaaag
 20 caaaaagtaaacacagtcacttctcaaaaaggatataagtttgaaacaattcaactagtt
 atgaatgaaatagatttttctcaagacgaagaacattagaccatttattgcaacgtgat
 ttagagaaagtcataataaaaattgtagaaaatatgacagtgataaaagtgttattaaa
 accatagaggcactcatgagaaaaggctataattatgataaaattaaatctaaattagaa
 gaaagcggatatctaatgaataa
 25

Sequence 882

MPKVTKIEVQKKNKERFNFLDGEFEMGIDIDTLVKFNKKDQILEPSDMQNIQEYDHYR
 RGVNLAIQYLSYKKRTEREVIQYLEKNDIQSNAIQDVIDYCYKEKFIHDHEDYAESLKNTM
 IHTTDKGPEIYRQKLYQLGIEVTIEKYVEAYEQQPLDDVIKVAEKVMKSKKGPEAKVK
 30 QKVTQSLQKGYKFETIQLVMNEIDFSQDEETLDHLLQRDLEKVNKNCRKYDSKSVIK
 TIEALMRKGYNYDKIKSKLEESGISNE*

Sequence 883

Contig_0524_pos_2369_2668,
 35 putative peptide of unknown function
 atgagtgaaatgtccgagcaagaactaagacatgaaatacaattatTTaaagaaaaaatg
 cgtaaagcagagatgaatggcattatgaatgaatatgatgtttatcaaagcaaagtgatt
 atagcagaaagctatcttgtgatcgcaataaaattgaacctggaaaaatctataaaactc
 aatgatggtagtaaacagtaactttaaagtagaacgactcaaggggtgatttgcattgggga
 40 tttagaataaacagtagtgaaacctgaggaaggtctaccattagcattattaaaatttttag

Sequence 884

MSEMSEQLRHEIQLFKEKMRKAEMNGIMNEYDVYQSKVIAESYLVDRNKIEPGKIYKL
 45 NDGSKQYFKVERLKGVFAGWFRINSSEPEEGLPLALLKF*

Sequence 885

Contig_0524_pos_2978_4201,
 putative peptide of unknown function
 50 gtgagccaatatgtgagcgaactcgtacaactttttccttatgaagtaactgagcataaa
 gttgaacaaattattcaatgggcacatttaggtgattacatagaagaaaaggtttctaat
 ttaagcgaaaaatcatacgcacaacttttacttagtattgcacgttcttcaaaaaacgat
 attatcatttttaaatcatgttttatcacatttagatgaaacatttatagaaagagctaca
 gttttatcaaaagatttatattgaagctaataaaacactagtcttaattgataatgatgta
 55 gaaaagataagcaaaaacaaagtaactacataacatgggtatcacacggccagatagcaag
 gaaggttctctaaatcaggtattaccaacttttagagaacatgaaaaagatcgtagt
 ttaaaatcagaaatggaaattgagaactttgattatgattggaagcagaatcgctctcgt
 attcctgaaatgacttataattttaaaagaatagaacgttataatcatgctaagccacca
 agatttctagtgagattttggaccttatttgcagctttttgattgggcttgttttaatg

agcgtacttttctttaacaacttaggtatgggtcaaactagggaaacattaatacccaagca
 tccatacagaatcagaataaagatacttacgaagaaaaacttgcatatggattagcatta
 gatggatcagttactttaaacgggtctaaagatttaaaagtacctaagtatagtttaatt
 5 acaattactggagaaaaataacaaaagatatcggtcgaaatgaatcaaaggagatatagt
 gttagtaaaaaatcaagtgttttatttcaatccagctgggttatacgaatctcactttt
 aagaaattgtcaccttatataaaaatcaaattatagtacttacgtagagtactttaacagt
 cacttacatcaaaaacatgataaagtaacagaaacgcttagacctgataaagataaaaaag
 tatgtgtaccgatcacgcaacaacctataaaaaatgatatttggtgataatgataaactg
 tctggatttgttattccaatgacaaataaaacggaattgaaaaaacatttaatatcacg
 10 aaagatgtatggattacaaaaagtggaagcgggtattttatcgctgatatgaaagaagaa
 aaatggatttatattgaattgtag

Sequence 886

VSQYVSELVQLFPYEVTEHKVEQIIQWAHLGDYIEEKVSNLSEKSYAQLLLSIARSSKND
 15 I I I L N H V L S H L D E T F I E R A T V L S K D Y I E A N K T L V L I D N D V E K I S K T S N Y I T W V S H G Q I R K
 E G S L N Q V L P T F R E H E K D R T S L K S E M E I E N F D Y D W K Q N R S R I P E M T Y N F K R I E R Y N H A K P P
 R F L V R E W T L F V S F L I G L V L M S V L F F N N L G M V K L G N I N T Q A S I Q N Q N K D T Y E E K L A Y G L A L
 D G S V T L N G S K D L K V P K Y S L I T I T G E N N K R Y R V E M N Q R R Y S V S K N Q V F Y F N P A G L Y E S H T F
 K K L S P Y I K S N Y S T Y V E Y F N S H L H Q K H D K V T E T L R P D K D K K Y V V P I T Q Q P I K M I F G D N D K L
 20 S G F V I P M T N K T E L K K T F N I T K D V W I T K S G S G Y F I A D M K E E K W I Y I E L *

Sequence 887

Contig_0525_pos_396_2015,
 is similar to (with p-value 0.0e+00)
 25 >sp:sp|P45554|DNAK_STAAU DNAK PROTEIN (HEAT SHOCK PROTEIN 70
) (HSP70). >gp:gp|D30690|STANHS_3 Staphylococcus aureus gene
 s for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds. NID:
 g487326.
 atgggtacagattataaagtagatattgaaggtaaatacatatacaccacaagaactttca
 30 gcaatgatattttacaaaattttaaaaagcactgcagaaaactatttaggggatcacgtagac
 aaagctgttatcactgtccctgcttatttcaatgatggtagaactgaagcaactaaagat
 gctggtaaaattgcaggcttagaagttgaacgtattatcaacgaacctacagctgctgca
 cttgcttatgggttagataaaaactgaaacagatcaaaaggctctcgtatttgacttaggt
 gggggaacatttgacgtatctattctagagtttaggcgacgcggtatttgagattatca
 35 actgccggagataataaaacttggtggcgatgacttcgaccaagtgattattgattatctt
 gtttcagaattcaagaagagaatgggtgtagatttatcacaagataaaaatggcattacaa
 agattaaaagatgctgccgaaaaagctaaaaaagatttatcagggtgtttctcaactcaa
 atttcattaccattcatttctgctggagaaaatggccattacacttagaaattagttta
 actcgttctaaaatttaggaatttagctgattcattaatcaaaaaactatggaaccgact
 40 cgtcaagcattataaagatgctgggttatctacttcagaaatagatgaagttattttagtt
 gggtggttcaacacgtattccggccggtcaagaagctgttaaaaaagaaaattgggaaagaa
 ccacataaagggtgttaaccagatgaagttgtagcaatgggtgctgctattcaagctggt
 gtaatcacaggtgatgttaaagatgtagtattacttgatgttacgccattatctttaggt
 atcgaatttatgggtggagctatgaacacattaattgaacgtaataactactattccaact
 45 tccaaatcacaagtttattctacagcagctgacaatcaaccagcagtagatatcatgta
 ttacaagggtgaacgtccaatggcatctgacaacaaaacttttaggaagattccaattaact
 gacattccacctgcaccacgtgggtgtacctcaaatcgaagtaacatttgatcgcataaa
 aacgggtattgttaacgttacagctaaagatttaggtactaataaagaacaaaacattaca
 atacaatcaagctcatctctatctgatgaagaatcgatcgcatgggtgaaagatgctgaa
 50 gaaaatgctgaagcagataaaaaacgtcgtgaagaagtagacttgcgaaacgaagcagat
 agtctagatttccaagttgaaaaaacagttaaagacttaggcgaaaatattagcgatgaa
 gataagaaaaatgctgaagagaaaaaagatgcacttaaacagcattagaaggtgaagac
 atcgacgatattaaagctaaaaaagaagaacttgaaaaagtaattcaggaattatctgca
 aaagtttatgaacaagctcaacaagcacaacaagaaggccaagaagaacaagggttctcaa
 55 gatagcactgttgaaagatgcagactttaaagaagttaaagatgacgaagataaaaaataa

Sequence 888

MGTDYKVDIEGKSYPQELSAMILQNLKSTAENYLGDTVOKAVITVPAYFNDGERQATKD

AGKIAGLEVERIINEPTAAALAYGLDKTETDQKVLVFDLGGGTFDVSILELGDGVFEVLS
TAGDNKLGDDFDQVVIDYLVSEFKKENGVDLSQDKMALQRLKDAAEKAKKDLSGVSQTQ
ISLPFISAGENGPLHLEISLRSKFEELADSLIKKTMETPRQALKDAGLSTSEIDEVILV
GGSTRIPAVQEA VKKEIGKEPHKGVNPDEVVAMGAAIQAGVITGDVKDVLVDVTPLSLG
5 IEIMGGRMNTLIERNTTIPTSKSQVYSTAADNQPAVDIHVLOGERPMASDNKTLGRFQLT
DIPPA PRGVPQIEVTFDIDKNGIVNVTA KDLG TNKEQNITIQSSSSLSDEEIDRMVKDAE
ENAEADKKRREEVDLRNEADSLVFQVEKTVKDLGENISDEDKKNAAEEKKALKTALEGED
IDDIKAKKEELEKVIQELSAKVYEQAQQAQQQGQEEQGSQDSTVEDADFKEVKDDEDEKK*

10

Sequence 889

Contig_0525_pos_2160_3281,

is similar to (with p-value 0.0e+00)

>sp:sp|P45555|DNAJ_STAAU DNAJ PROTEIN (HSP40). >gp:gp|D30690
15 |STANHS_4 Staphylococcus aureus genes for ORF37; HSP20; HSP7
0; HSP40; ORF35, complete cds. NID: g487326.

atggccaaaagagactattatgaagtccttaggcgtaaacaaaagcgcttctaagacgaa
attaaaaaagcttatcgtaaattatcaaaaaataccatccagatataaataaagaagaa
ggcgagacgaaaaattcaaaagaaatctccgaagcatatgaagttttaagtgatgaaaac
20 aaactgtgcaattatgatcaatttggatcatgacggaccacaaggcggaatttggagtc
gaacttgggtggcagtgactttgggtgatttgaagatatttccagctcattcttgggtggc
ggttcacgtcaaaagagatcctaatagcacctcgcaaaaggtgatgacctcaatacacaatg
acaataacatttgaagaggctgtattcgggacaaaaaagaaatatcaataaaaaaagat
gtaacatgtcatatcatgaacgggtgatggggctaaacctgggtacaagtaaaaaaattgt
25 agctattgtaatggcgctggctgttctgttgaacaaaatactatttgggtagagt
agaactgaacaagtttgtcctaataatgtgaaggtagtggacaagaattgaagaacctgt
ccaacatgtataaaggaaaaggactgaaaataaaaacagttaaactagaagtaactgttcct
gaagggtgtagataacgaacaacaagttcggttagctggagaagggtcacctgggtgtaac
ggaggaccacatgggtgacctatgtgggtgttcagagttaaacctccaatacatttgaa
30 cgtgatggagacgatattactataatctagatattagcttttcacaggctgcactaggt
gatgaaattaagatacctacattaaaaagtaattgtgttttaaccattccggcaggtaca
caaacgggtaaacaaattccgacttaaaagataaaaggtgtaaagaatgttcatggttatggc
tacggggacttattgtcaacataaaagtggttacaccaacaaaattaaatgaccgtcaa
aaagaattattaaaagaatttgcgtgaaatgaatggtgaaatatataatgaacagtcatt
35 aatttcaagatagagcgaaaagattctttaaaggagaatag

Sequence 890

MAKRDYYEVLGVNKSASKDEIKKAYRKL SKKYHPDINKEEGADEKFK EISEAYEVLSDEN
KRANYDQFGHDGPQGGFGSQGFGGSDFGGFEDIFSSFFGGGSRQRDPNAPRKGGDLQYTM
40 TITEEAVFGTKKEISIKKDVTCHTCNGDGA KP GTSKKNCSYCNAGRVSV EQNTILGRV
RTQVCPCKEGSGQEFEEPCPTCKGKTENKTVKLEVTVP EGV DNEQQVRLAGEGSPGVN
GGPHGDLVYVFRVKPSNTFERDGGDIYNNLDISFSQAALGDEIKIPTLKSNNVLTIPAGT
QTGKQFRLKDKGVKNVHGYGYGDLFVNIKVVTPTKLNDRQKELLKEFAEINGENINEQSS
NFKDRAKRFFKGE*

45

Sequence 891

Contig_0525_pos_3285_3893,

is similar to (with p-value 7.0e-83)

>sp:sp|P45557|PRMA_STAAU PROBABLE METHYLTRANSFERASE (EC 2.1.
50 1.-). >gp:gp|D30690|STANHS_5 Staphylococcus aureus genes for
ORF37; HSP20; HSP70; HSP40; ORF35, complete cds. NID: g4873
26.

atgaattggatggaactctcaattgtagtttaatcacgaagtagaatacagatgttacagaa
attcttgaaagtattggctctaattggagttgttaattgaagattcaaatattttagaagaa
55 caacctattgataaagtttggagaaatttatgacttaaacctgaagactatcctgaaaaa
ggagttcgattaaaagcttactttaatgagttcacttataatgaaaacttaaaatccaac
atcaattatgaaatattaagtcttcagcaaatgataaaacaatttatgattaccaggaa
aaacttattgccgaagtagattgggaaaatgaatggaagaattattttcatccatttaga
gcttcaaaacaatttacgatagtagcaagttgggaatcatatgttaaagaaaatgataac

gaattgtgcattgaattagatccaggtatggcttttggaaacaggtgatcatccaacgaca
agtatgtgttttaaagcaattgaaacttttgtaaaaccaactgattcagttatcgacgtt
ggaacagggtcaggcattttaagtattgctagtcatttacttggagttcaaagaataagg
gggatttga

5

Sequence 892

MNWMELSIVVNHEVEYDVTEILES YGSNGV VIEDSNILEEQPIDKFGEIYDLNPEDYPEK
GVR LKAYFNEFTYNENL KSNINYEILSLQQIDKTIYDYQEKLIAEVDWENEWKNYFHPFR
ASKQFTIVPSWESYVKENDNELCIELDPGMAFGTGDHPTTSMCLKAIETFKPTDSVIDV
GTGSGILSIASHLLGVQRIRGI*

10

Sequence 893

Contig_0525_pos_1838_1506,

is similar to (with p-value 5.0e-42)

15

>sp:sp|P45554|DNAK_STAAU DNAK PROTEIN (HEAT SHOCK PROTEIN 70
) (HSP70). >gp:gp|D30690|STANHS_3 Staphylococcus aureus gene
s for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds. NID:
g487326.

20

atgtcttcaccttctaatgctgttttaagtgcattctttttctcttcagcatttttctta
tcttcacgcctaataatcttcgcctaagtctttaactgtttttcaacttgaataactaga
ctatctgcttcgtttcgcaagtctacttcttcacgacgtttttatctgcttcagcattt
tcttcagcattctttcaccatgcgacgatttcttcacgacgatttcttcagcattt
attgtaattgtttgttctttatttagtacctaaatcttttagctgtaacgttaacaataccg
tttttatcgatatcaaatgttacttctgatttga

25

Sequence 894

MSSPSNAVLASAFFSSAFFLSSSLIFSPKSLTVFSTWNTRLASFRKSTSSRRFLSASAF
SSASFTMRISISSDRDELDCIVMFCSLLPKSLAVTLTIPFLSISNVTSI*

30

Sequence 895

Contig_0526_pos_555_1499,

is similar to (with p-value 0.0e+00)

35

>gp:gp|U15783|SEU15783_1 Staphylococcus epidermidis orf334 p
rotein, putative multidrug resistance protein QacC, and QacC
' genes, complete cds. NID: g622953.

40

atgacgaaaagtgggaacaacgcccatggagagaaaagaagatagataatgtaagttat
gcagatatactggaaattttaaaaaataaaaaaggcttttaattgtaaaacaattgtggaac
gtcttagagttcaagccgactgatgaaggttatttgaagttacataagacatggttttgt
aagtcgaaactctgccagtttgaattggaggcgtgctatgaaaaatagttatcaagct
caaaaagtgttgaagaagttgtttaaagaaaaaaccaaaagcgcgttggttatttttaaca
ctttcaacgaaaaatgcgatatagatggggatactttagaacaaagtttgaacatttaacg
aaagcatttgataggttaagtagatataaaaaagtgaagcaaaatcttgttggtttttg
cgttcaacggaagtaacagtttaataaaaatgatggttagttataatcaacatatgcatgtt
ttattatgtgttgaataatagttattttaagaataaagctaattatataactcaagaagaa
tggttgaatttatggcaaaaagcattacaagtaaatatcgaccgtagcaaatattaaa
gcgatcaaaccaaatcaaaaaggcgataaagatatcaagcagctatcaagaaacctct
aaatattcgggttaagtcacgtgattttttaactgatgatgaaagaaatcaagaatc
gtgaatgatttagaaaaagggtttatatcgaaaacgtatgttgagttatggtgggttgctt
aaacaaaaacataagattttaaattagatgatgccgaagatggcaatttgattaataca
agtgacgaagataaaacaacagacgaagaagaaaaagcacattcaattacggcaatttg
aattttgaaaaacaaaattattattttaaaagatttgaaacgtag

50

Sequence 896

MTKSGKQRPWREKKIDNVSYADILEILKIKKAFNVKQCGNVLEFKPTDEGYLKLHKTWFC
KSKLCPVCNWRRAMKNSYQAQKVIEEVVKEKPKARWFLTLSTKNAIDGDTLEQSLKHLT
KAFDRLSRYKKVKQNLVGFRLRSTEVTVNKNDSYNQHMVLLCVENSYFKNKANYITQEE
WNLWQKALQVNYRVPANIKAIKPNQKGDKDIQAAIKETSKYSVKSSDFLTDDDERNQEI
VNDLEKGLYRKRLSYGGLLKQKHKILNLDDAEDGNLINTSDEDKTTDEEEKAHSAITAIW
NFEKQNYLYKDLKR*

55

Sequence 897

Contig_0526_pos_2971_3645,
is similar to (with p-value 0.0e+00)

- 5 >sp:sp|P14506|TRA1_STAAU TRANSPOSASE FOR INSERTION SEQUENCE
ELEMENT IS257 IN TRANSPOSON TN4003. >pir:pir|S04162|S04162 t
ransposase 1 - Staphylococcus aureus plasmid pSK1 transposon
Tn4003 >gp:gp|X13290|SATN4003_1 Staphylococcus aureus multi
-resistance plasmid pSK1 DNA containing transposon Tn4003. N
10 ID: g46747. >gp:gp|X13290|SATN4003_6 Staphylococcus aureus m
ulti-resistance plasmid pSK1 DNA containing transposon Tn400
3. NID: g46747. >gp:gp|U40259|SEU40259_11 Staphylococcus epid
ermidis trimethoprim resistance plasmid pSK639. NID: g176207
9. >gp:gp|U40381|SEU40381_1 Staphylococcus epidermidis plasmid
15 pSK697 insertion sequence IS257(697A) putative transposase
gene, complete cds. NID: g1762091. >gp:gp|U40384|SEU40384_1
Staphylococcus epidermidis plasmid pSK818 insertion sequence
IS257(818A) putative transposase gene, complete cds. NID: g
1762097. >gp:gp|AF051916|AF051916_1 Staphylococcus aureus pl
asmid pJE1 remnant of replication protein Rep (rep), trimeth
20 oprim resistance protein DfrA (dfrA), thymidylate synthetase
ThyE (thyE), and putative transposase Tnp (tnp) genes, comp
lete cds; and unknown gene. NID: g3676404. >gp:gp|AF051916|A
F051916_7 Staphylococcus aureus plasmid pJE1 remnant of repl
25 ication protein Rep (rep), trimethoprim resistance protein D
frA (dfrA), thymidylate synthetase ThyE (thyE), and putative
transposase Tnp (tnp) genes, complete cds; and unknown gene
. NID: g3676404. >gp:gp|AF051917|AF051917_21 Staphylococcus
aureus plasmid pSK41, complete sequence. NID: g3676412.
30 atgaactatttcagatataaacaatttaacaaggatgttatcactgtagccgttggtac
tatctaagatatgcattgagttatcgtgatatactgaaatattaaggggacgtgggtga
aacgttcatcattcaacggtctaccgttgggttcaagaatatgccccaatatttatcaaa
atttggaaagaaaagcataaaaaagcttattacaaatggcgtattgatgagacgtacac
aaaataaaaaggaaaatggagctatttatatcgtgccattgatgcagagggacatacata
35 gatatttgggttgcgtaagcaacgagataatcattcagcatatgcgtttattaacgtctc
attaacaatttggtaaacctcaaaaggtaattacagatcaggcaccttcaacgaaggta
gcaatggctaaagtaattaaagcttttaaaccttaaacctgactgccattgtacatcgaaa
tatctgaataaacctcattgagcaagatcacgcgtcatattaaagtaagaaagacaaggat
caaagtatcaatacagcaagaatacttttaaaaggattgaatgtatttacgctctatat
40 aaaaagaaccgcaggtctcttcagatctacggattttcgccatgccacgaaattagcatc
atgctagcaagtttaa

Sequence 898

- MNYFRYKQFNKDVITVAVGYLYRLYALS YRDI SEILRGRGVNVHSTVYRWVQEYAPILYQ
45 IWKKKHKKAYYKWRIDETIYIKIGKWSYLYRAIDAEGHTLDIWLKQRDNHSAYAFIKRL
IKQFGKPQKVITDQAPSTKVAMAKVIKAFKLKPDCHCTSKYLNNLIEQDHRHIKVRKTRY
QSINTAKNTLKGIECIYALYKKNRRSLQIYGFSPCHEISIMLAS*

Sequence 899

- 50 Contig_0526_pos_3744_4484,
is similar to (with p-value 3.0e-23)
>sp:sp|P30267|YKAA_BACFI HYPOTHETICAL 50.9 KD PROTEIN IN KAT
A 3'REGION (ORF A). >pir:pir|S27491|S27491 hypothetical prot
ein A - Bacillus firmus >gp:gp|L02548|BACKATA2_1 B.firmus OR
55 F A and ORF B, complete cds. NID: g143118.
gtgccggtaacgattaataataaactcttcaattttatttagatcattttgtcacatggata
agtagcgcattacctcttttaactaagatattcataatgattatcattatactaggtgct
atttatccattttattaaagggacatggaatcggaataccgttgaaacaatttttagttta
tttaaagttttgggagtcatttataggcgttttggttaatttttaacattgggccaagtgg

ttacttaatgaacaaacgggaatgtatgtttttaactatttggttaattccggtaggatta
 acagtacctgcaggagcgcggtattagctttattagtaggatatggcttattagaattt
 gtaggtgtttatgcgcaaaaaattatgtaccgatatggaaaacgcctggacgttcagca
 gttaatgcttttagcatcttttgttgctagttttgctgtgggttacttataacgaataaa
 5 gagtataaagaaggtaaattcacggaaaaacaagctgttatcatagcaaccggcttttct
 acagttactgtagcttttatgatagttattgctaaaaccttacacttaattggatatatgg
 aatttatatttttggctctaccttgtttgttactgctgcagtaaacagcttgtacagttagg
 atttggcctatcagtaaaattagcaacacatattatgatcagccatttatagaagaagat
 acaagcgaattaaaaggttaa

10

Sequence 900

VPVTINNNSILLDHFVTWISSALPLLTKIFIMIIILGAIYFFIKGTWNRNTVETIFSL
 FKVLGVIIIGVLLIFNIGPSWLLNEQTGMVFNLYLVPVGLTVPAGGAVLALLVGYGLLEF
 VGVYAQKIMYPKWTPGRSAVNALASFVASFVGLLITNKEYKEGKFTKQAVIIATGFS
 15 TVTVAFMIVIAKTLHLMDIWNLYFWSTLFVTA AVTACTVRIWPIKISNTYYDQPFIEED
 TSELKG*

Sequence 901

Contig_0528_pos_686_1450,
 20 is similar to (with p-value 5.0e-56)
 >sp:sp|P39605|YWCG_BACSU HYPOTHETICAL 28.3 KD PROTEIN IN QOX
 D-VPR INTERGENIC REGION. >pir:pir|S39698|S39698 hypothetical
 protein - Bacillus subtilis >gp:gp|X73124|BSGENR_44 B.subti
 lis genomic region (325 to 333). NID: g413923. >gp:gp|Z99123
 25 |BSUB0020_106 Bacillus subtilis complete genome (section 20
 of 21): from 3798401 to 4010550. NID: g2636240.
 gtgggtatagtgtcagattatgtttatgagttgatgaacaacatcattcagttagaaaa
 tttagaatcaaccacttggttctgaaacggtagaaaaattagtagagcgcgacagagt
 gcttctacatccagttatcttcaaacttattctattattggtggtgaagatccaagcatt
 30 aaagcgcggtttaaaggaagtgtcaggtcagccttatgttttagataatggttattttatt
 gtatttgttttagattattatcgctcatcatttagtagatgaagttgcgcgctcaaatatg
 gagacatcatatggttctgcagaaggactatttagtaggtacaatagatgttgcattagtt
 gcgcaaaacatggcagttgctgcccgaagatatggggtatggaattgtttatttaggtca
 ttgcgtaatgatgttgcgcgagtgcgtagaattttaaatttacctgattatacgtttccg
 35 ttatttggtagtggtcagtagtgaaaccttctgatgaagaaaatgggtcacctaaaccgcg
 ttgccatttaaacatatttttcataaaagaccagtatgatgcgaatcagcatcaacaacgt
 aaagaattggaagcatacgaccaagtagtgagtgaatattataaagaacgtactcacggt
 gtgcgtacagaaaattggtcacaacaaatagaacatttctaggacgtaaaacacgttta
 gatatgttagatgaattgaaaaaagcaggatttattcaagataa

40

Sequence 902

VGIVSDYVYELMKQHHSVRKFNQPLGSETVEKLVEAGQSASTSSYLQTYSIIGVEDPSI
 KARLKEVSGQPYVLDNGYLFVFLDYRHHLVDEVAASNMENTSYSAGSAGLLVGTIDVALV
 AQNMVAEAEDMGYIGIVYLGSLRNDVARVREILNLPDYTFPLFGMAVGEPSEENGSPKPR
 45 LPFKHIFHKDQYDANQHQQRKELEYDQVVSEYYKERTHGVRTENWSQQIETFLGRKTRL
 DMLDELKKAGFIQR*

Sequence 903

Contig_0528_pos_2809_3198,
 50 putative peptide of unknown function
 atgaggtgtaataacaatgccaaatcagatacctattgctaaagcggtaaagacacgttta
 gggaatgagacgtgttttcttgccataatatttaacattattaggaaaaaagtaacacg
 atgacattaattactgtaaaaagtatttccatatacgatatatagctcctttttaattttt
 agtattccgacatatgtaataagaataaagggggttaagaagaaattcaagttgttgatt
 55 tatagtttccagattcatttaagttgaatagagtgaaggtgattactagaacaatgaaa
 aatgagaagtttatgtttatcagaaaagtgagcaaaagaggacgtaaaagaggtttatatta
 ataaaaaagacgcgtgcaatcatgctataa

Sequence 904

MRCNTMPNTIPIAKAVKTRLGNETCFLAIIFNIIRKTSNTMTLITVKSISISISYSSFLIF
SIPTYVIRIKGVKKKFKLLIYSFQIHLKLNVRKVITRTMKNEKFMFIRKVSKEVDKRFIL
IKKTRAIML*

5 Sequence 905

Contig_0528_pos_2962_1574,

is similar to (with p-value 0.0e+00)

>sp:sp|P54596|YHCL_BACSU HYPOTHETICAL 49.0 KD PROTEIN IN CSP
B-GLPP INTERGENIC REGION. >gp:gp|X96983|BS75DGREG_13 B.subti
lis chromosomal DNA (region 75 degrees: cspB upstream of glp
PFKD operon). NID: g1239975. >gp:gp|Z99108|BSUB0005_181 Baci
llus subtilis complete genome (section 5 of 21): from 802821
to 1011250. NID: g2633055.

atggaaatactttttacagtaattaatgtcatcgtgttacttggttttcctaataatgtaa
15 aatattatggcaagaaaacacgtctcattccctaaacgtgtctttaccgcttttagcaata
ggatcgtatttggcattgtattacacctcatatatgggtgcagagtctaaaactctcgaa
caatcaacagactggtttagtattgttggagatggttatgttgactattacaaatgatt
gtcatgccactaatattcattcaattgttgccgcttttagcaaaaatacaaatgggtgaa
aaattcgctaagatcggttcttataatttttatgtttttaattggtactgtagccattgca
20 gctatcgttggaaattttttacgctttgatctttggttttagatgcacgtctattgattta
ggtagtgcagaacattcacgtggtacagaaatttcaaaacaagccaaagatttaactgca
aacactttaccacaacaaattctcgaagtattcccaagcaatccatttttagatttcaca
ggacaacgtacaacttcgacaattgcagttgttatttttgcaacgtttgtgggctttgct
tatcttagagttgcaagaaaacagccggaacatggaagcttacttaaacgtggtatagaa
25 gcaatctattctatcggttatggctatcgtaactttgttttacgattaacgccttatggc
attttagctattatggcttctactcttgcgacaagtatttttctgcaatttggtacgtta
ggtaaatcttaattgcttcatacgcagctctaatacacaatgtatattatccatttaatt
atactgagtgcttaggtatcaatcccggttaaatacgtgaaaaagacaatagaagtacta
atctttgcatttacttcacgttcaagtgcaggtgcattaccggttaattgttcaaacgcaa
30 acaaaacgttttaggtgtacctgagggaaattgcaaaacttctctgcaacttttggtttatcc
atagggcaaaatggctgtgcaggaatctatcctgctatgctagcagttatgggtggcacca
gtagcaaatgtagaaattgacttccaatttgttgttacacttattgctgttattata
agttcatttggcgttgcaggtgtaggtggcggggcaacattcgcatcaatactcgtatta
tctacacttaactaccagttgctctcgaggggtactgatattctatcgaaacctctcatc
35 gatatgggtcgtacagcacttaacgttaatgactcaatgctagctggaacaggtaccgca
cgcttaacgaatcattgggacaaaaaacatttgactcaaatgattacggcgatttatct
gcaaatata

Sequence 906

40 MEILFTVINIVILLVFLIMLNIMARKHVSFPRVFTALAIGIVFGIVLHLIYGAESKLE
QSTDWFSIVGDGYVALLQMIVMPLIFISIVAAFSKIQIGEFKAKIGSYIFMFLIGTVIAIA
AIVGIFYALIFGLDASSIDLGSAEHSRGTEISKQAKDLTANTLPQQILEVFPSPFLDFT
GQRTTSTIAVVI FAFVGFAYLRVARKQPEHGSLLKRGIEAIYSIVMAIVTFVLRLLTPYG
ILAIMASTLATSDFSAIWTLGKFLIASYAALITMYIIHLIILSVLGINPVKYVKKTIEVL
45 IFAFTSRSSAGALPLNVQTQTKRLGVPEGIANFSAFGLSIGQNGCAGIYPAMLAVMVAP
VANVEIDFQFVVTLIAVVISSFGVAGVGGGATFASILVLTSLNLPVALAGVLISIEPLI
DMGRTALNVNDSMLAGTGARLTNHWDKKTFDSNDYGDLSAN*

Sequence 907

50 Contig_0528_pos_1216_866,

is similar to (with p-value 1.0e-22)

>sp:sp|P39605|YWCG_BACSU HYPOTHETICAL 28.3 KD PROTEIN IN QOX
D-VPR INTERGENIC REGION. >pir:pir|S39698|S39698 hypothetical
protein - Bacillus subtilis >gp:gp|X73124|BSGENR_44 B.subti
lis genomic region (325 to 333). NID: g413923. >gp:gp|Z99123
55 |BSUB0020_106 Bacillus subtilis complete genome (section 20
of 21): from 3798401 to 4010550. NID: g2636240.
gtgacccattttcttcatcagaagggttcacctaactgccataccaaataacggaaacgtat
aatcaggtaaatatttaaaatttcacgcactcgcaacatcattacgcaatgaccctaatt

aaacaattccataccccatatcttcggcagcaactgccatgttttgcgcaactaatgcaa
catctattgtacctactaatagtccttctgcagaaccatatgatgtctccatatttgacg
ccgcaacttcatctactaaatgatgacgataataatctaaaacaaatacaataaataac
cattatctaaaacataaggtgacctgacacttcctttaaacgcgctttaa

5

Sequence 908

VTHFLHQVHLLPYQITETYNQVNLKFHALAQHHYAMTLNKQFHTPYLRQQLPCFAQLMQ
HLLYLLIVLLQNHMSPYLTPLHLLNDDNNLKQIQINNHYLKHKADLTPLNAL*

10

Sequence 909

Contig_0528_pos_0_556,

putative peptide of unknown function

atgacactaacaataaagaagttgctaaagttttatttaaagcttatagatataaaaaa
cccatcgatttccattagtgagaactatcaattaaacgaagaagaagcatatcatgtacaa
gaagaactaattgaccaattaactttcaaagaccgttcgactgttacagggtataaagtt
agtatgactagcaaggcaacgaagcaattgctaactaacgaacctgcatatggaaca
ctcttatctaaccaaattgttaatgatggcgctcagtcctctcttcagaattattttca
ccattactagaaccagaaattatctttatagtcgaggaagacttaccttatgatgctgat
ttagaacaattagatatcataccggtatcgcgccaggcattgaaattccagatgcaaga
tataaaaattggtttccaaattttacttttatcagatttaatatcagataataaccgcaaca
ggacttgcgtagtaggtgaccctgtagacggacttgataacgatgcatttgctaattgta
catttaaatTT

15

20

Sequence 910

25

MTLTNKEVAKVLFKAYRYKKPIDFISENYQLNEEEAYHVQEELIDQLTFKDRSTVTGYKV
SMTSKATQAIANTNEPAYGTLLSNQIVNDGASVLSLSELFSPLEPEIIFIVQEDLPYDAD
LETIRYHTRIAPGIEIPDARYKNWFPNFTLSDLISDNTATGLVVVGDPVDGLDNDAFANV
HLNLYX

30

Sequence 911

Contig_0530_pos_5055_4645,

is similar to (with p-value 6.0e-48)

>pir:pir|S39743|S39743 hypothetical protein - Bacillus subti
lis

35

atggaccctaaagtagctatgttaagcttttctacaaaaggttctgctaaatcggatgat
gttactaaagtgaagaagcattgaagtttagctcaagaaaaagctgaagcagatcaatta
gatcatgtagttattgatggagaattccaatttgacgctgctattgttcctagcgtagca
gagaagaaagcacctgggtgcaaaaattcaaggtgatgcaaatgtatttgtttccctagt
ctagaagcaggtaatattggttataagattgctcaacgttttagtggtatcgatgcagta
ggaccagtcctacaaggattaaactctccagtcattgatttatctcggtgttgctcaact
gaagacgtttataacttatctattattacagctgctcaagctttacaataa

40

Sequence 912

45

MDPKVAMLSFSTKGSAKSDDVTKVQEALKLAQEKAEDQLDHVVIDGEFQFDAAIVPSVA
EKKAPGAKIQGDANVFVFPFSLEAGNIGYKIAQRLGGYDAVG PVLQGLNSPVNDLSRGCST
EDVYNLSIITAAQALQ*

Sequence 913

Contig_0530_pos_4573_3806,

is similar to (with p-value 5.0e-66)

>sp:sp|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KD PROTEIN IN PTA
3'REGION. >pir:pir|S39745|S39745 hypothetical protein - Bac
illus subtilis >gp:gp|X73124|BSGENR_91 B.subtilis genomic re
gion (325 to 333). NID: g413923. >gp:gp|Z99123|BSUB0020_60 B
acillus subtilis complete genome (section 20 of 21): from 37
98401 to 4010550. NID: g2636240.

55

atgcaatcttttgcgtttgatgacactttttccgaaagcgttggttaaagatttatcttgt
aatgtagtacgaacgttgatacatcaacacaccgtgattttgggcattcatgattcgct
ttaccatttttaagtgtggtattcggttttcttacagatgaacaaggatataatgcaatt

gtaggaattctggtggcttgggtgtcgtattagatcaaggaattttaaacatatctttg
 atttttaaaggacaaaccgaaacgactattgatgaagcctttacagtgatgtatttattg
 attaataaaatgtttgaggatgaagatgttagtatcgataactaaagaaattgagcaatcg
 tattgcccaggaaaaatttgatttaagtattaatgataagaaatttgccgggatttcgcag
 5 cgacgagtacgtggtggtatcgagtgcaaatatacttatgtattgaaggttctggctca
 gaacgggcattaatgatgcaacagttttatcaacgtgcgcttaaaggggagactactaaa
 tttcactatccagacatagatccctcatgtatggcatctttagaaacccttttaaataga
 gaaattaaagtgaagatgttatgtttttattattatgcactaaaagatttaggggca
 aacttaaatatggatcctattacagaagacgagtggacacgttacgaagggtattatgat
 10 aagatgttagaacgcaatgcgaaaatgaatgaaaaattagatttttag

Sequence 914

MQSFAFDDTFSESVGKDLSCNVVRTWIHQHTVILGIHDSRLPFLSDGIRFLTDEQGYNAI
 VRNSGGLGVVLDQGILNISLIFKQGTETTIDEAFTVMYLLINKMFEDEDVSI DTKEIEQS
 15 YCPGKF DLSINDKKFAGISQRRVRGGIAVQIYLCIEGSGSERALMMQQFYQ RALKGETTK
 FHYPDIDPSCMASLETLLNREIKVQDVMFLLLYALKDLGANLNM DPITEDEWTRYEGYYD
 KMLERNAKMNEKLDF*

Sequence 915

20 Contig_0530_pos_3054_2179,
 putative peptide of unknown function
 atgggtgaacacgcagttacatttgggtcaaccggcaatcgcaattccatttaattgctgga
 aaaattaaagtcctcattgaaagtttagatgaaggttaattattcttctatcacaagtgc
 gtatatgacggaatgttatagcatgccccgaacatctaaagtctatcattaatcgcttt
 25 gttgaaaaaagtggaagtgaagaaccactatcagtaaaaattcaaaactaatttgcctcca
 tcaagaggttttaggttcaagtgtcagtagcagtagcgtttgtagcgccagttatgat
 tttatggatcaaccttttagatgacaaaacattgattaaagaagcaaatgggcgagcaa
 atcgacatggttaagccaagcggatttgatacgacagcagattgtgtcaaataaaccgctc
 tggtttaaaacaagggcagggccgaaaaattaaaatcactaaaattaaatggttatatggtt
 30 gtcattgatactggagtaaaaggttctaccaaaacagcagtagaagatgttcatgtatta
 tgtgaatctgatgaatatatgaaatatatagagcacattggtacacttgttcacagtgct
 agcgaatcgattgaacagcatgatttccatcatttggctgacataatttaacgcagtgcaa
 gaagacttgagacatttaacagtaagtcacgataaaaatagaaaaattacttcaaattggg
 aaagaacatggtgccattgctggttaactaactggtggaggaagaggtggcagcatgctt
 35 ctctctgcggaatttaaaaaactgcaaagactattgttgctgctggtgaaaaagctggc
 gcagcacatacatggattgaacatttaggaggttaa

Sequence 916

MGEHAVTFGQPAIAIPFNAGKIKVLIESLDEGNYSSITSDVYDGMLYDAPEHLKSIINRF
 40 VEKSGVKEPLSVKIQTNLPPSRGLGSSAAVAVAFVRASYDFMDQPLDDKTLIKEANWAEQ
 IAHGKPSGIDTQTIVSNKPVWFKQGQAEKLSLKLNGYMVVIDTGVKGSTKQAVEDVHVL
 CESDEYMKYIEHIGTLVHSASESIEQHDFHHLADIFNACQEDLRHLTVSHDKIEKLLQIG
 KEHGAIAGKLTGGGRGGSMLLLAENLKTAKTIVA AVEKAGAAHTWIEHLGG*

45 Sequence 917

Contig_0530_pos_1178_102,
 putative peptide of unknown function
 atgattcaggtaaaaagcccccgaaaaactttatattgcaggcgagtatgcagtaaccgaa
 ccaggatataaatctattcttattgcagtaaatcgctttgtaacggcgacaattgaggcg
 50 tcaaataaagttgaaggttagtattcattccaaaacattacattatgaaccagttaaattt
 gaccgtaatgaagatagaattgaaatctcagatgttcaagctgctaagcaactgaaatat
 gttgtgacagctatagaagtgtttgaacagtatgtgcgcagttgcaatatgaatttaaag
 cactttcatttaaccattgatagtaacttagcagataactctggtcagaagtacggatta
 ggttcaagcgccgctgttttagtatctgttggttaaaagctttgaatgaattctatggttg
 55 gaattatcaaaacctttatatttataaaattagctgtaattgcaaataatgaaattacaaagt
 ttaagttcatgtggcgatattgcggttagtgtctacagtggttggttgcatatagtacg
 tttgaccatgactgggtgaaacagcaaatggaagaacatcggtgaatgatgttttgaa
 aaaaattggccagggttacatatcgaaacctttacaagctcccgaataatggaagtcctt
 attggatggactgggtctccagcttcttctccacacttagtgagtgggtcaaacgttta

5 aaatcagatccaagtttttatgggtgatttttagatcaatctcatgcttgtgtagaaagt
 ttaatccaagctttttaaactaataatatcaaaggtgttcaaaagatgatacgtataaac
 agacgtattattcaatctatggataacgaagcatcagttgaaattgaaacagataagcta
 aaaaaattatgtgatgtcgggtgaaaagcacggtggcgcttctaaaacttcaggtgctggt
 ggtggcgattgtggcattactattatcaacaaggaattgataaaaatattatttataac
 gaatggcaaatgaatgatatacaaccattgaaatttaaaatttatcacgggcaataa

Sequence 918

10 MIQVKAPGKLYIAGEYAVTEPGYKSILIAVNRFVTATIEASNKVEGSIHSKTLHYEPVKF
 DRNEDRIEISDVQAAKQLKYVVTAIEVFEEQYVRSCNMNLKHFHLTIDSNLADNSGQKYGL
 GSSAAVLVSUVKALNEFYGLELSNLYIYKLAVIANMKLQSLSSCGDIAVSIVSGWLAYST
 FDHDWVKQMEETSVNDVLEKNWPGLHIEPLQAPENMEVLIGWTGSPASSPHLVSEVKRL
 KSDPSFYGDFLDQSHACVESLIQAFKTNNIKGVQKMIRINRRIQSMDSNEASVEIETDKL
 KKLCDVGEKHGGASKTSGAGGGDCGITIINKVIDKNIYNEWQMNDIKPLKFKIYHGQ*

Sequence 919

Contig_0531_pos_1619_3223,
 is similar to (with p-value 5.0e-67)

20 >sp:sp|P54417|OPUD_BACSU GLYCINE BETAIN TRANSPORTER OPUD. >
 gp:gp|AF008220|AF008220_90 Bacillus subtilis rrnB-dnaB genom
 ic region. NID: g2293135. >gp:gp|Z99119|BSUB0016_80 Bacillus
 subtilis complete genome (section 16 of 21): from 2997771 t
 o 3213410. NID: g2635411. >gp:gp|U50082|BSU50082_1 Bacillus
 subtilis glycine betaine transporter OpuD (opuD) gene, compl
 25 ete cds. NID: g1524396.
 atggactggagcacttcataaggcgtagtcattgtgttacttttctgctgttatacctatg
 atggtttttccgaaagcaagtgaataatcattaccgatatacaatagtgccatttcta
 tcaattggatcggtatatactctttatgggactggctatattttgtttgtttatacata
 gcatttggtaagtatgggaatgtcacggttaggaaaagcgactgacaaacctgaattta
 30 aatttcacatgggcagccatgttattctgtgccggtattggttcagatattttatattgg
 ggtgttattgagtgggcattttattatcaagtacctcctaacggtgcaaaatcaatgtcc
 gatcaagcacttcaatatgcaactcaatatggtatgtttcactggggacctatagcctgg
 gcaatatatgtgctaccagctttgccaatcggtattttagttttcggttaagaagaaccc
 gtctataaaattagtcaagcttgcgaccaattttaaaaggacatacggataaattatta
 35 ggaaaaatcgtagatattttattttttcggtttgctcggtggtgctgcaacatcactc
 gctctaggcgtgcccgatgatctcagctggtattgaacgattgactgggttagatggatct
 aatatgattttacgttcaatcatcttactaactattacagttattttcgcaatcagttct
 tacacaggtttgaaaaaaggatttcaaaaattaaagtgatgttaacggtttggttatcatt
 ttattattagcatttgcattcatcgtaggtccaactgtgtttattatggaaactacagtt
 40 acagggttcggtaatatgataaaagatttcttccatatggcgacatggatggaaccattt
 ggtggcataaaaggctcgtaaagaaacgaatttccctcaagattggacaatattctactgg
 tcatggtggctcggtttatgcaccgtttattggattgtttatcgcgcgatctcaaaagga
 cgtacacttaaagaagttgtattaggaacaatatgctatggaacattaggttgtgtgta
 tttttcggtatttttggttaactatgctgtatatctacaaattactgagcaatttaagtga
 45 ataagctatttaaacaattatggtacagaggcaacaatcatagaaataatgcatcaacta
 ccattctcgacaattactattatcttattcttaatatcagctttcttattcttagcaaca
 acattcgattctggttcataatatttagcagcagcgctcacagaaaaaagtgataggagaa
 ccgttacgtgctaactggtttgtctggcggtttgcgttatgtttactaccgttctcttta
 atgctagttggaggagaacgtgcattagaagtattgaaaacagcatcattacttgctagt
 50 gtacctttaattgttatattacgctaattgatgatttcgttcttaattatactcgacga
 gatcgtatcaagtttagaaagacgtgcagataagcataaagaaattgaaagacgttctcta
 agaatagttcagggtcaaagacaaacctgaagacgataacttataa

Sequence 920

55 MDWTTTFIGVVIVLLFAVIPMMVFPKASEIIITDINSAINSIGSVYLFMGLAIFCFVLYI
 AFGKYGNVTLGKATDKPEFNNETWAAMLFCAGIGSDILYWGVIEWAFYYQVPPNGAKSMS
 DQALQYATQYGMFHWGPIAWAIYVLPALPIGYLVFVKKKPVYKISQACRPILKGHTDKLL
 GKIVDILFIFGLLGGAATSLALGVPMISAGIERLTGLDGSNMILRSIILLTITVIFAIS
 YTLKKGIQKLSNVNVLWLSFLLAFVFIIVGPTVFIMETTVTGFGNMIDKDFHFMATWMEPF

GGIKGRKETNFPQDWTIFYWSWWLVYAPFIGLFARISKGRTLKEVVLGTICYGTLGCVL
 FFGIFGNYAVYLQITEQFNVISYLNNGTEATIIIEIMHQLPFSTITIILFLISAFLLAT
 TFDSGSYILAAASQKKVIGEPLRANRLFWAFALCLLPFSLMLVGGERALEVLKTASLLAS
 VPLIVIFTLMMISFLIILGRDRIKLERRADKHKEIERRSLRIVQVKDKPEDDNL*

5

Sequence 921

Contig_0531_pos_3494_3811,
 is similar to (with p-value 3.0e-30)

>gp:gp|Z99119|BSUB0016_180 Bacillus subtilis complete genome
 (section 16 of 21): from 2997771 to 3213410. NID: g2635411.

>gp:gp|U47861|BSU47861_2 Bacillus subtilis gbsAB operon, gl
 ycine betaine aldehyde dehydrogenase GbsA, alcohol dehydroge
 nase GbsB genes, complete cds. NID: g1524391.

gtgtcaatttcacgttcccatTTTTtagtaaaaaagttgcggaagaaagtgaagaaatct
 15 tttctctgaataaaatgttgcttccggtaccacgcgtgaattgtgtttgacgatatcg
 tattctttagtTTTTtacacctgtactcatactaggtttactcatttgacgttgtgtg
 cgcatttcatacaagtgtcatacttccttcaaaaaccataatgccatacaagttacctaca
 ctacggttgataccatacaaatccatggtttcaccgattgagttgataactaaatcttta
 gttcttctcgatatattga

20

Sequence 922

VSISRSHFLVKKLRKKVKSFSIAIKCCFRLPRVNCCLTISYSCSFFTPVLILGLLICSCC
 RISSSVILPSKTIMPYKLPTRLRIPYKSMVSPIELITKSLASSIY*

25

Sequence 923

Contig_0531_pos_3650_3276,
 is similar to (with p-value 6.0e-26)

>gp:gp|Z99119|BSUB0016_180 Bacillus subtilis complete genome
 (section 16 of 21): from 2997771 to 3213410. NID: g2635411.

>gp:gp|U47861|BSU47861_2 Bacillus subtilis gbsAB operon, gl
 ycine betaine aldehyde dehydrogenase GbsA, alcohol dehydroge
 nase GbsB genes, complete cds. NID: g1524391.

atgagtacaggtgtataaaaaactacaagaatacgaatcgatcgtaaaacaattcacgcgt
 ggtagccgaaagcaaacattttattgcagagaaagatttcttcactttcttcgcaacttt
 35 tttactaaaaaatgggaacgtgaaattgacactaatatggaagcaattgaagatgcagaa
 aatatcattcaccacttcttgaaaagaatgatatagatgaggaagttaaaaaacaagca
 ataaatgtaaaagcgcagtttagaccactctaaaatatattacaagtggtcgctcaatta
 agtgaagctctagagtcaggagaaatttttaactattttcccaattccagatgaacaacat
 catcatgatcaataa

40

Sequence 924

MSTGVKKLQEQYDIVKQQFTRGSRKQHFIAEKDFFTFFRNFFTKKWEREIDTNMEAIEDAE
 NIIHPLLEKNDIDEEVKKQAINVKAQLDHSKIYYKWLALQLSEALESGEIFNYFPPIPDEQH
 HHDQ*

45

Sequence 925

Contig_0531_pos_1883_1578,
 putative peptide of unknown function

atggctgcccattgtgaaattattaaattcaggtttgtcagtcgcttttcttaacgtgaca
 50 ttcccatattaccaaattgctatgtataaaacaaaacaaatatagccagtcaccataaag
 agatataccgatccaattgaattagaaatggcactattgatatcgtaattgattatttca
 cttgctttcggaacaaaccatcataggtataacagcaaaaagtaacacaatgactacgcct
 atgaagtcgtccagtcaccataacttttcttttcaaattgtgatccccctaattataat
 ttatga

55

Sequence 926

MAAHVKLLNSGLSVAFPNVTFPYLPNAMYKTKQNIASPIKRYTDPIELEMALLISVMIIS
 LAFGKTIIGITAKSNTMTTPMKVVQSITFSFFKCDPPNINL*

Sequence 927

Contig_0532_pos_1470_709,

is similar to (with p-value 4.0e-40)

- >sp:sp|P54717|YFIA_BACSU HYPOTHETICAL 29.3 KD PROTEIN IN GLV
 5 G-GLVBC INTERGENIC REGION. >gp:gp|Z99108|BSUB0005_88 Bacillu
 s subtilis complete genome (section 5 of 21): from 802821 to
 1011250. NID: g2633055. >gp:gp|D50543|D50543_2 Bacillus sub
 tilis DNA for 76-degree region, complete cds. NID: g1486240.
 10 atgatttttagatgaacgtgtaaaactctaatttcgatcaattaaatgataatgatatacaa
 attgcacattatgttaatacacatatagatgtttgcaaaaatatgaaaatacaagattta
 gcctcacagacacatgcttcaaatgctacgattcatcgcttcactcgtaaactaggtttt
 gatggttatagtgaactttaaatcctttttaaaatttgaagatagtaagaatcatcaactt
 ccttctgattctatggagcaatttaacaagaattgaaaatacattcaactatttagaa
 cgatttgattatcgctttatttaactcacaaaatgcatcatgctacaacaatacttatat
 15 ggtactggacgtgcacagatgaatgtcgctgaagaagcacaacgtatactgttgactatg
 cataaaaaatattatattgttacatgatgttcataactaaagatgggtgttaacaagaca
 attccagaagatttgtttttcatcatttcactttctggcgaaacacatcaacttaaagaa
 gtcacacaattgcttcaactgagacaaaaatattttatttccgtaacaacaatgaaagac
 aatacattggcacaacaagctgattacaatgtctatgtttcaagcaataccttctattta
 20 aacgatggactgattattccagttttatttagctatcacattttctttgaaacactacta
 agaaaatataacgaatataaagagaatcatgaattaacatag

Sequence 928

- MILDERVNSNFQNLNDNDIQIAHYVNTHIDVCKNMKIQDLASQTHASNATIHFRTRKLG
 25 DGYSDFKSFLKFEDSKNHQLPSDSMEQFKQEIENTFNYLERIDYRLLTHKMHHATTIYLY
 GTGRAQMNVAEEAQRILLTMHKNIILLHDVHELKMLVNLKTIPEDLFFIISLSGETHQLKE
 VTQLQLRQKYFISVTTMKDNTLAQQADYNVYVSSNTFYLNDGTDYSSFISYHIFETLL
 RKYNEYKENHELT*

30 Sequence 929

Contig_0533_pos_907_1239,

putative peptide of unknown function

- gtgattaggaatagccctccgataacaaatgaaatagacagagggaaattcatagtgaca
 tgtaaagcgattgctcctaaaaatccatgcaattcccgctcctattgttaataattcttata
 35 acagcttttagaaatgccttttaattctctaaaaatctagattactactaccttcaaataat
 ataattgctacagcaagagatacaattgaactaaatgcctcaggtccaagtgcctctttt
 ggatttgctaataccaaaaataggtcctacaagtaaacctacgatggccatgacaacaatc
 gatggccattttattctactcgctaaccattga

40 Sequence 930

VIRNSPPIITNEIDRGKFIVTCKAIAPKIHAIPAPIVILITALEMPFNSLKSRLLLPSNN
 IIATARDTIELNASGPSASFANPKIGPTSKPTMAMTTIDGHFILLANH*

Sequence 931

- 45 Contig_0533_pos_1649_2698,

is similar to (with p-value 1.0e-16)

- >sp:sp|P37520|YYAD_BACSU HYPOTHETICAL 37.7 KD PROTEIN IN RPS
 F-SPO0J INTERGENIC REGION. >pir:pir|S18084|S18084 hypothetic
 al protein 9 - Bacillus subtilis >gp:gp|D26185|BAC180K_52 B.
 50 subtilis DNA, 180 kilobase region of replication origin. NI
 D: g467326. >gp:gp|X62539|BSORIGS_14 B.subtilis genes rpmH,
 rnpA, 50kd, gidA and gidB. NID: g40020. >gp:gp|Z99124|BSUB00
 21_199 Bacillus subtilis complete genome (section 21 of 21):
 from 3999281 to 4214814. NID: g2636442.
 55 atagcaatgttaagaatccatcattgttgcttcgctttgtagcgcttgtagttgga
 gcaggatttgcgacaggtcaggaaattttcaattcttcactagtcattgttattacagt
 attggcggtatttttattactggacttatcataactcttgagggaatattcgtattaaat
 actggctttcgtcttagatctcaaaaccactctgaatctattcgttattatttacatcca
 acaatagctaaaattatttgatattatacttacagtatattttattttcttagcaattatt

atgacagccggcgagcatcgactataaatgaaagttttggcttaccttttggttaagt
 tctctcatttttagtgatacttattttgttaacattatttctaaagttcgatcgacttatc
 gctgttttaggaggggtaacaccatttcttgtggcagtcgtagtaattgattgcagtgat
 tactttattaccggtgatttaacttttagtgacgtcagtcgaatttcaaatacaaaataag
 5 tgcatttcacctgggttggtggtttgacgcaattaattatgctagcttacaattgctgct
 gcatttagctttttaactgtaattggcggttaagctacgatataaatcgccacaatttat
 ggtggacttatcggtggtattattgtgactttactattacttttgattaattttggtcct
 gttacagaatttaatacaattaaagaggttagcattaccatcattgctacttgctaagcag
 atttctccatccattggtattatcatgtctgtcattatggttttagtcatatacaataca
 10 gtagtaggttaattgtacgcctttgcatcacgcttttagtcgaccgtttacgaaacgctat
 tatattcttatagttatgatggcaataataacatttgcttgacttttgtgggattcatt
 tctctcattggtaagtggtccctattatgggactttttggttttatcttattgattcct
 gtgatatacaaaaggaattttacgaaaataa

15 Sequence 932

MSNVKESIIVAFVGVVVGAGFATGQEIFQFFTSHGIIYSIGGIFITGLIITLGGIFVLN
 TGFRLRSQNHSESIRYYLHPTIAKLFDIILTVELFLSLAIIMTAGGASTINESFGLPFWLS
 SLILVILILLTLFLKFDRLIAVLGGVTPFLVAVVVMIAVYFITGDLNFSQSVSNQNK
 SISPGWFFDAINYASLQIAAASFSLTVMGGKLRYSSTIYGGIGGIIVTLLLLLINFLGL
 20 VTEFNQIKEVALPSLLLAKQISPSIGIIMSVIMVLIYNTVVGLMYAFASRFSRPFCKRY
 YILIVMMAIITFACTFVGFISLIGKVFPIMGLFGFILLIPVIYKGILRK*

Sequence 933

Contig_0533_pos_6099_6896,
 25 is similar to (with p-value 2.0e-38)
 >sp:sp|P54721|YFIE_BACSU HYPOTHETICAL 31.5 KD PROTEIN IN GLV
 BC 3'REGION. >gp:gp|Z99108|BSUB0005_93 Bacillus subtilis com
 plete genome (section 5 of 21): from 802821 to 1011250. NID:
 g2633055. >gp:gp|D50543|D50543_7 Bacillus subtilis DNA for
 30 76-degree region, complete cds. NID: gl486240.
 atgacaaattttcattctatagatgctacacaagtaacaaacgtcactttaaatgttaaa
 gatttaataaatttaactgatttctattctaatgtattaggttttctattcaaaaacaa
 acgaatcaacaaaccgtattcaacatcggaatcttggttatactttaactttaaatgaa
 ctttaacaacggtcgacaaccggaatttagagaagcagggttattccatgttgcttatctt
 35 ttaccgactcgtagcgatttagcagacttcctttatcatgctaacaatctcaacatcgca
 atgggtggtggagatcaccttgtagtgagcgctatatttctactgatcctgaagcgaat
 ggtattgaagtctatcatgatcgcccttcagaagactggttggtggcgagacggtttgtc
 aaaatggatacattggaagtgtgatgtcaatgatttaatactcaacggtcaaatgaaggt
 tggcaaatgttgccggaagaaggaataatcgggcatttacatctcaaaacacacaattta
 40 gaatctgcttatgaattttatggtgaaaagttagggttcgaacatatatctaatttccca
 caagcactatttatgtccactcaaaagtatcatcatcatatagctacaaatacttggcag
 tcaataaagattagaactcaaaatgaacaaacttatggtttatgtcactttgacatatat
 caacctaatagcaataactactcatgtttacctcacctgaaggctttgacattacaattcat
 45 ggtaacgaaacaaaataa

Sequence 934

MTNFHSIDATQVTNVTNLNVKDLNKLTDYFSNVLGFSIQKQTNQQTTFNIGNLGYTLTLNE
 LNNGRQPEFREAGLFHVAYLLPTRSDLADFLYHANNLNIAAMGGGDHLVSEALYFTDPEGN
 GIEVYHDPSEDWLWRDGFVKMDTLEVDVNDLMTQRSNEGWSWPEEGKIGHLHLKTHNL
 50 ESAYEFYVEKLGFEHISNFPQALFMSTQKYHHHIIATNTWQSNKIRTQNEQTYGLCHFIDIY
 QPNANTTHVTSPEGFDTIHGNETK*

Sequence 935

Contig_0533_pos_5371_4442,
 55 is similar to (with p-value 3.0e-24)
 >gp:gp|Z71552|SPADCA_4 Streptococcus pneumoniae adRCBA oper
 on. NID: g3758891.
 gtgtcaatcatcatttttaattgtaagcggatgcagtagctttgatcatcgtaaacgcgaa
 agtattaatgacaagaataaaatgaaagtatacacgactgtatatgcatttcaaaagtttg

acacaacagattggtggaaaatatgttgacgcgcaatcaatctatcctgctggtgctgat
 ttacactcatatgaaccaacacaaaaagatatgattgatattgccaaaagtgatctgttt
 gtctattcaagtcattcaattagatcctgtcgctgcaaagattacgaattcgatgaccaat
 aatagcatgaaattagcgcttgccgaaggactcaaaacaaagtgattttattcactctaaa
 5 gaccatgatgaaaatcatgagcatcatcaccatcatgaagaatcgaatcaagatcctcat
 gtttggttagatcctgtttctaaatcaaaaattcgctttcatgattaaagagaaattaata
 gagaaagaccctaaacatcaagcttattacaataaaaattataaaatagtaataaagat
 attgtgcatattgatcaacaactacaatcaataacgaagcattctaaaagagataaagtt
 gtgatcaccacgattcgcttggtatatttagcgcatcggttatggttttaacaacaaggt
 10 gttaaaggtatgaatgatgaagaacctagtcaaaaagagattttgaatatcgtaaagat
 atacagcattcacatgcgccttatgttttatatgaacaaaattattacctccaaaattaca
 gatgttattaagaagaacagatacgaaccattaagttttcataatctagctgtattg
 actaaaaaggagcaaaatgatgattcaatttcataccaatcattaatgaaaaagaacatt
 tacgtattaaatcgcgactcaataattaa

15

Sequence 936

VSIIILMLSGCSSFDHRKRESINDKNMKVYTTVYAFQSLTQQIGGKYVDAQSIYPAGAD
 LHSYEPTQKMDIAKSDLFVYSSHQLDPVAAKITNSMTNNSMKLALAEGLKQSDFIHSK
 DHDENHEHSHHEESNQDPHVWLDPLVNQKFAFMIKEKLEKDPKHQAYYNKNYKIVNKD
 20 IVHIDQQLQSITKHSKRDKVVISHDSLGYLAHRYGFKQQGVKGMNDEEPSQKEILNIVKD
 IQHSHAPYVLYEQNITSKITDVIKKETDTKPLSFHNLAVLTKKEQNDDSIYSQSLMKKNI
 YVLNRLNN*

Sequence 937

25 Contig_0534_pos_5367_5684,
 is similar to (with p-value 5.0e-23)
 >gp:gp|U56999|TPU56999_1 Treponema pallidum methyl-accepting
 chemotaxis protein (mcp-1) gene, complete cds, and potentia
 l regulatory molecule (pfoS/R) gene, partial cds. NID: g1354
 30 774.
 atggctacagttgacaatggtgatattgattataaaactaaacaccatagagattaacata
 ctcataaacactgggttgtaattcagtaaaagctattaaccatgtcaccaattcccgttgta
 attaaccttacatattggttaaagtaaaatcgccgattgtagctgctaaaccaccaacaact
 gtcggaaaaacaattaaagccatactaccaacacgcttcttctattaataaaataagtaac
 35 acggcgatggatgcagttaacatcggtattaattaaatccccaattcccgaattacccat
 gtaccttggtttaaattga

Sequence 938

MATVDNGDMIKLNITIEINILINTGCNSVKLLTMSPIPVVINLTYGKVNTPIVAAKPPTT
 40 VGKTIKAILPTRSSINKISNTAMDAVNIVLIKSPIPAITHVPCLN*

Sequence 939

Contig_0534_pos_8940_0,
 is similar to (with p-value 1.0e-30)
 45 >sp:sp|P25468|PYRD_SALTY DIHYDROOROTATE DEHYDROGENASE (EC 1.
 3.3.1) (DIHYDROOROTATE OXIDASE) (DHODEHASE). >gp:gp|X55636|S
 TPYRDD_1 Salmonella typhimurium pyrD gene for dihydroorotat
 e dehydrogenase (EC 1.3.3.1.). NID: g854623.
 atgtacaaattagttgaagcctttattattcaaattagatcctgaacgagcacatggtttg
 50 accatcaatgcgttggaagtgtgttcaaaaatgttcacccattttacctatcgtaataag
 ttatttacttataacaatccaatattaacgcaacacattcaccggtatttcttttgataat
 cctatcggttgatcgaggttttgataaatcttggaagttccaaaagcacttgaaaac
 attggcttcgggtgcaattgaaactcggcggtataacacctaagcctcaaccaggtaatcca
 aaaccacgcatgtatcgtttactagaagatgatgcactcatcaatcgatgggattcaat
 55 aataagggtatgaataaagcactaagtaatttacgtaatcattcatgctcaataaccagta
 ggattaaatggttggtgtgaataaaaacaacttcctatgaaaatcgctatcaagattaca

Sequence 940

MYKLKPLLFKLDPERAHLTINALKCVQKCSPIPIVKNKLFYNNPILTQHIGISFDN

PIGLAAGFDKSCEVPKALENIGFGAIELGGITPKPQPGNPKPRMYRLEDDALINRMGFN
NKGMMNKALSNLRNHSCSIPVGLNVGVNKTTSYENRYQDYX

Sequence 941

5 Contig_0534_pos_8398_7859,
putative peptide of unknown function
atgcatagccaataccaacaagataataaatttggattccgtttaccacatgaaggtgca
gatatttcctttgataattcatggactgagacatggaaagagattttataaatcgtaga
atggatcacttacaagatgagttattacgtgtaggattgtggaaacaagaagataaaaaa
10 atgtatgaacgtgtaagaaaagttattgttgatgaactttcaaatacactagtaagccc
tctctgttacatgggtgatttatggggaggtaactacatgttcttaacaaatggccaacct
gctttatttgatcctgcaccactatatggagatagagaatttgacataggaatcactaca
gtatttgggtggatttacacaagagttctatgatgaatataatcaacagttaccactagcc
aagggatcacaaaagcgtatagaattttatagattatatttacttatgatacatttactt
15 aaatttgggtggtatgtatgctgatagtgtaacaacgctctatgaaaatcattttagaataa

Sequence 942

20 MHSQYQQDNKFGFRLPHEGADISFDNSWTETWKEIFINRRMDHLQDELLRVGLWKQEDKK
MYERVRKVIIVDELSNHTSKPSLLHGLDWGGNYMFLTNGQPALFDPAPLYGDREFDIGITT
VFGGFTQEFYDEYNQQLPLAKGSQKRIEFYRLYLLMIHLLKFGGMYADSVQRSMKIILE*

Sequence 943

25 Contig_0534_pos_5971_4913,
is similar to (with p-value 6.0e-49)
>gp:gp|U56999|TPU56999_1 Treponema pallidum methyl-accepting
chemotaxis protein (mcp-1) gene, complete cds, and potentia
l regulatory molecule (pfoS/R) gene, partial cds. NID: gl354
30 774.
atgtcgacgataaaaaatatcgatggaccaaaggattttgttttttagagtgttatcaggg
gtagcaattggaatagtagccggactcggtccaaatgcaattttgggagaaatttttaaa
tactttatgcaatatcatcctattttcaaaaactttattaggggtcggtcaagccatccaa
tttacagtgccagcgcttattggagcattgatagctatgaagttcaatatgacaccttta
35 gcaatagctgtagtagcaagtgctcatatgttggtagtggtgcagctcaatttaaaaca
ggtagcatgggtaattgcggaattggggatttaattaatacagatgttaactgcacccatc
gccgtgttacttattttattaataagaagaacgtgttggtagtagtggcttaattgttttt
ccgacagttgttggtggttttagcagctacaatcgcgctatttactttaccatatgtaagg
ttaattacaacgggaattggtagcatgggttaataagctttactgaattacaaccagtggtt
40 atgagtagtgaatctctatggtgttttagttttataatcatatcaccattgtcaactgta
gccatagctattgctattggattatcaggtattgctgcgggatctgctcaataggtata
gcagcgacagaagctgtattattgattggtagcagcaagttaatcatgtaggtattcct
ttatcaatatttttcgggtgggtgaaaatgatgatgcaaatatggttaaataccctgtc
attatgattccgattttcttgacagcggcaatatctggtattgcttcagggattattggt
45 atttcaggaacaaaagaatcagcaggatttgggtttatcggaatggttgggcctattaat
gcctttaaatttatgcatgttgattctgcatggttaagtttattacttattgtcatcgcc
ttttttgtgtgcccgtttctagttgcatggatttttagatttaataacttagaagattaatt
catttgtagagaatgatatttttaaatattatgggataa

Sequence 944

50 MSTIKNIDGPKDFVFRVLSGVAIGIVAGLVPNAILGEIFKYFMQYHPIFKTLLGVVQAIQ
FTVPALIGALIAMKFNMTPLAIAVVASASYVGSAAQFKQGTWVIAGIGDLINTMLTASI
AVLLILLIEERVSGMALIVFPTVVGGLAATIGVFTLPYVRLITTGIGDMVNSFTELQPVF
MSMLISMVFSFIIISPLSTVAIAIAIGLSGIAAGSASIGIAATEAVLLIGTSKVNHVGP
55 LSIFFGVKKMMMPNMVKYPVIMIPILTAASIGIASGIIIGISGTESAGFGFIGMVGPIN
AFKFMHVDSAWLSLLLVIAFFVVPFLVAWILDILRLIHLIYENDIFKFMG*

Sequence 945

Contig_0534_pos_4812_3388,

is similar to (with p-value 4.0e-37)

>gp:gp|U12891|PAU12891_5 *Pseudomonas aeruginosa* PAO substrain OT684 pyoverdine gene transcriptional regulator PvdS (pvdS) gene, complete cds. NID: g1580798.

5 atgtcagggaaactagaagaattacaattaaaagtagctcgattaagtcgacgtactcat
gaattaggtattccaattatggtattatttgaggggattcctgcttcggggaagacacgt
ttatcaaataaatactattgcacctagatgccaaatattcgcgatttatagctactaaa
tcgccagagtcacacgatttacgttaccaatttttatacaaaaatattggaatactttacca
caaaagggaataataaattttattttagaagttggtattcacactttttagattataaa
10 gaaaataaaataagcatgatcaatataaaaattatgatgttttagtcaatcaaatttat
cattttgaatcgatgttaagaatgataactatgaaattataaaaattttcatagaaata
aatgaagaaaaacgcaatgaacatattcaacagacaaaagataatccattaactagatgg
aaagttcaagaatgatgaaatgttatacctcaagaaagttatctaaatcaaatacatcaa
ttcatcaacaaagataaagattggaaagtgatcgattacacagagcgcgagcatgctttt
15 gaaaaaatgtacttacatttaatatagatagacttgagcaagctataaaaaaagttgaacaa
caacaactaaagtcaacggtaagttcacatcaagctttacgacttctttatttaataat
aatcttgagaaagtagacaaaaaacgtataaaaatctcattgttgtaattgcaacagaga
atgagagaaatccaatttgctttatatgaaagaaagattccccctgttttggttttcgaa
ggatggtgctgctggtgaaaggtggcaatattaaacgtattagagaaaaattagatcca
20 acaggatatgaagtgaatggtattagtgacactacggatgtcgaacttaagcatcattat
ttgtggagatttgctaaaaagatgccaaaatcaggtcatatagaaatatttgatcggagt
tggtatggtcggtgactagttgaacgtgtagaaggttttgcaagccagaatgaatggcaa
cgagcatctgatgaaatcaatcaatttgaaaagatgtggacagatgaaggtacaatcata
ttaaaattcttcttatgttttagataaagatgagcagcttaagcgttttaagaccgtgaa
25 aataatcctgataaacaatggaagattactgaagaagattggcgtaatagagaaaaatgg
gatgaatatttagaagcaagtcgatgatgattgaatctacaaacacttcatatgccctt
tggtatattgttccggcagatcataaaaaaacgagtcggattgaagtacttaaaacaatt
attagaaaaatgtgaagaagtactatggggagtttaagacgtattaa

30 Sequence 946

MSGKLEELQLKVARLSRRTHELGIPIMVLFEGIPASGKTRLSNELLLHLDKYSRFIATK
SPESNDLRYQLQKYWNTLPQKGNINIYFRSWYSHFLDYKENKIKHDQYKNYDVLVNQIY
HFESMLKNDNYEIIKFFIEINEEKRNEHIQQTNDPLTRWVKVEYENVIPQESYLNQMHQ
FINKDKDKWKVIDYTEREHAFKMYLHLIDRLEQAIKKVEQQTTKVNGKFTSSFTTSLFNN
35 NLEKVDKTKYKNLIVELQQRMEIQFALYERKIPLVLVFEGMDAAGKGGNIKRIREKLDP
TGVEVNGISAPTDVELKHHYLWRFKMKPKSGHIEIFDRSWYGRVLVERVEGFASQNEWQ
RASDEINQFEKMWTDGTTIILKFFLCLDKDEQLKRFKDRENNPDQWKITEEDWRNREKW
DEYLEASHDMIESTNTSYAPWYIVPADHKKTSRIEVLKTIIRKCEEVLWGKTY*

40 Sequence 947

Contig_0534_pos_2732_2025,
is similar to (with p-value 9.0e-23)

>gp:gp|U96108|SCU96108_5 *Staphylococcus carnosus* (3R)-hydroxymyristoyl acyl carrier protein dehydrase homolog (fabZ) gene, partial cds, YwpF homolog, single-strand binding protein homolog (ssb), SceD precursor (sceD), SceA precursor (sceA) and SceE precursor (sceE) genes, complete cds, and TenA homolog (tenA) gene, partial cds. NID: g2735509.

atgaaaaaacagttatcgcttctacattagcagtatcttttaggaattgcaggttacggt
50 ttatcaggacatgaagcacacgcttcagaaactacaaacggttgataaagcacacttagta
gatttagcacacataatcctgaagaattaaatgctaaaccagttcaagctggtgcttac
gatattcatttcgtagacaatggataccaatacaacttcacttcaaatgggttctgaatgg
tcatggagctacgctgtagctggttcagatgctgattacacagaatcatcatcaaaccaa
gaagtaagtgcacatacaatctagtaacacaaatgtacaagctgtttcagctccaact
55 tcttcagaaagtcgtagctacagcacatcaactacttcatactcagcaccaagccataac
tacagctctcacagtagttcagtaagattatcaaatggtaatactgctggttctgtaggt
tcatatgctgctgctcaaatggctgcacgtactggtgtatctgcttcaacatgggaacac
atcattgctagagaatcaaatgggtcaattacatgcacgtaatgcttcaggtgctgctgga
ttattccaaactatgccaggttgggggttcaactggttcagtaaatgatcaaatcaatgcc

gcttataaagcatataaagcacaaggtttatctgcttggggatgtaa

Sequence 948

MKKTVIASLAVSLGIAGYGLSGHEAHASETTNVDKAHLVDLAQHNPEELNAKPVQAGAY
 5 DIHFVDNGYQYNFTSNGSEWSYAVAGSDADYTESSNQEVSAANTQSSNTNVQAVSAPT
 SSESRSYSTSTTSYSAPSHNYSSSSSVRLSNGNTAGSVGSYAAAQMAARTGVSASTWEH
 IIARESNGQLHARNASGAAGLFQTMPGWGSTGSVNDQINAAKAYKAQGLSAWGM*

Sequence 949

10 Contig_0534_pos_0_1250,
 is similar to (with p-value 1.0e-51)
 >gp:gp|U93874|BSU93874_12 Bacillus subtilis cysteine synthas
 e (yrhA), cystathionine gamma-lyase (yrhB), YrhC (yrhC), Yrh
 D (yrhD), formate dehydrogenase chain A (yrhE), YrhF (yrhF),
 15 formate dehydrogenase (yrhG), YrhH (yrhH), regulatory prote
 in (yrhI), cytochrome P450 102 (yrhJ), YrhK (yrhK), hypothet
 ical protein YrhL (yrhL), putative anti-SigV factor (yrhM),
 RNA polymerase sigma factor SigV (sigV) and YrhO (yrhO) gene
 s, complete cds, and YrhP (yrhP) gene, partial cds. NID: gl9
 20 34604. >gp:gp|Z99117|BSUB0014_194 Bacillus subtilis complete
 genome (section 14 of 21): from 2599451 to 2812870. NID: g2
 634966.
 atgatgtttgtgtcacggttgttttaatatatacattactatttaaaccgaattaatt
 attagtataaaacatgatgctatagcagcacttttctatgtttcttaactggtggtatct
 25 atacaagcgtagactatttttaatacaatttgctgtagcgccattaaagcatttgtggtca
 ttagctattgaagaacaattctatctattctttccatttatacttttaggcttattaag
 tttttcaaaaagagaactacaatgattattctattaatcatctctttattatcattaact
 gcaatgataaacgatacatatgtatacaggttaacaattctagagtttatttcgggactgac
 acacgtttcaaacggttattattaggatgcttactagcatttatttggccaccgttctct
 30 ttcagaaaggatataatctaaaggtgctaaagcaagtataagtgaataggcatagtcgga
 atggcagtgctcatttatttggtagtggttagtgatgaagataaatggatatatagt
 ggaggattttatgccatctcgttcttaacactgtttgtcattgcaagcggttgtgcatcca
 tcaagtgttttaagaaaataactaagtttcaagttatttatttataggttaagagatcg
 tatagtttatatttattggcactatcctatcattatttttatgaatagttatttcgtacaa
 35 ggtcagattccttgggttggatataatttgtgaagtgatacttatgtttgtcatggctgaa
 gtatcttataaatttatcgaaacacctattagaaaaaatggatttaaagcattcacggtg
 ataccgaaaaatttaacaagattttcaagaacgattattgtgttaactcttgcttctcct
 tctgcattcatagattttggtgcctatgatagtttgggtaaaagagcatgataaacaacaa
 gctgcgaacaaaaatcttttaaaacgaaccagaaagcaaacctaaaaagccagatgaa
 40 aataatcaagataagttctcacaacaacattttaatcctaagaagcggtctccattattg
 ctgggagattcagtaatggtagatatcggtcaagtccttagtgaaaaagtacaaatgct
 aatattgatggaaaagttggccgacagttaattgagggtaaagatttaataatcaataaaag
 taccaagattatactaaaaaaggtcagagtggtgtgatagaacttggtac

Sequence 950

45 MMFVVTVVLIIYTLFLFKPELIISIKHDAIAALFYVSNWYIIQDVDYFNQFAVAPLKLHWS
 LAIEEQFYLFPPFILLGLLKFFKKRTTMIILLIISLLSLTAMITIHMYTGNNSRVYFGTD
 TRLQTLGCLLAFIWPFFSFRKDISKGAKASISAIGIVGMAVLIYLFVVVSDQDKWIYS
 GGFYAISEFLTLFVIASVHPSVVLKKILSFKLFIYIGKRSYSLYLWHYPIIIFMNSYFVQ
 50 GQIPWFVYICEVILMFVMAEVSYKFIETPIRKNKGFAFTVIPKNLTRFSRTIIVLILLVP
 SAFIVFGAYDSLKGHEKQQAQKQSFKTNQKAKPKPDENNQDKSSQQHFNPKAESPLL
 LGDSVMVDIGQVFSEKVPANIDGKVGRLIEGKDLINQKYQDYTKKGQSVVIELGT

Sequence 951

55 Contig_0535_pos_3501_3932,
 putative peptide of unknown function
 atgaagaatatggtaattctgaataagcaaaaaaggatgatcagaatgaaaaaagcaata
 ttttagtattattttctcttatttttagttcttaactgctactggatgtagtaataagttct
 aaagaaaaaccaattaaaaaagtgcatagaaattaatcctacaagtaagctgttaatt

attacagtaataaaaaagaaaataacaaacctgaaaaaattgggaaagtgtatcgatat
 aaaaaataacaatgcaaaagaaattactaacgacgggtattaaaaaagataactaaagataca
 ttgatttggaaaggtgtagcaacaaatcagataatgtaaaagatttattaggagaaaagt
 attctttatgaagttaaatataaaaaatggggatataaaaaaattcgagagaaaaattaaa
 5 tatactgaataa

Sequence 952

MKNMVILNKQKRMIRMKKAIFSIIISLILVLTATGCSNSSKEKPIKKSALINPTSKAVN
 ITVNKKENKPEKIGKVYRYKNNNAKEITNDGIKKDTKDTLIWKGVANKYDNVKDLLGES
 10 ILYEVKYKNGDIKKFERKIKYTE*

Sequence 953

Contig_0535_pos_5601_6350,

is similar to (with p-value 2.0e-17)

15 >gp:gp|X13481|BTPGI2XX_4 Bacillus thuringiensis plasmid pGI2
 with transposon Tn4430. NID: g3171732.
 atgaaaaagttatgatgagcttattatgttgacaaatattgtggaacagttttaactaca
 ggaagtgtaaatgcacaagggtgatagctcaactaactctgagtcgttaaaggaacttcaa
 aatgaaggtattgtgtctaatcaattactgaacaacaatggcaacaaatgaaagctcaa
 20 gagcgtaaagatgaagcagaatttgagaaaacagctgaagtacaatggcagaaacaacaa
 aaacaagatcgaatagatcgagaaaatcggttcaagaaaaagaaatttcatttgaaaaaa
 ggtgatttttcatcacaaataacgtaagttcaaaagggttacaggtcatgctgcaata
 tatactggaaaaggaaaagttaaagaagcgcttgatatggacaacctgtgagggtaaaa
 agtttttagtgattggaaaaagagtactttgaaaaaaagaaaaggaagtcctcaaacatcgt
 25 tatatcaagggtttatcgagcaccaaaaaatatagaggtaagctggaaactatgcgaaa
 tctcattttaacgggtgtaccttacagtataacgacgaatccatattctaaaagcgttaca
 tactgttcgaaacttgtttgcaatcgtattattatgggtgccggacgattatcagttctt
 ccggttggttacatctcaatttattattgagccatatagtcctcaataaatatattccatca
 aaagcagtttaggtcttataaaaagaagctaa
 30

Sequence 954

MKKLMMSLLCCTICGTVLTTGVSNAQGDSSSTNSESLELQNEGIVSKSITEQQWQQMKAQ
 ERKDEAEFEKTAEVQWQKQKQDRIDRENRSRKKKFKHLKKGDFFITNNVSSKGFTGHAAI
 YTGKGVKEAPGYQPVVRVKSFSWKKSTLKKRKGSPKHRYIKVYRAPKKYRGKAGNYAK
 35 SHFNGVPYSITTPNPKSVTYCSKLVQSYYYGAGRLSVLPVVTSSQFIIEPYSLNKYIPS
 KAVRSYKRS*

Sequence 955

Contig_0535_pos_7096_7773,

40 putative peptide of unknown function

gtgggcatttttagtatcggggtcagggtatagcgagtgtaacaaacaatatataactcacgca
 aaagaaagtcacgattcaactcctcaaaaatttaaatttagtgggaacgtatgatacttct
 caagttgattccaaaacgatgaaacaatttaaagaaatagaaaaagaagataataatttc
 cacataactaaacatggaaataaagtcgtttagaagacaaattacctaattccagagaat
 45 aaaacttcaagttattcagctgatggtagtgctgaaaataatacaaaaagtaatttaatttc
 tctgattttgttggaatatggatgggaaagatgatggaaaaatatcggtgggataacc
 ttttatagtggttaaatcatataacggacaacacgatgggtcaaaaagtaaaaaagggact
 catgtacattgtaatagatttaacggaacaaaatctgatcatagatactgggtcaaaaaaa
 catcctagagcttatgtagattttataaaaagtgttgctgggtatcacgccaagcttat
 50 aaatgttcttcttgggaaaaatgactaaatgcgatggtttgaatagtatttatagaaaa
 ggtgtcaaagattgctcatcatggaaaggtaaacccaaacataaaaaactggcctaaaaaca
 gcatggtatagaaattaa

Sequence 956

55 VGILVSGSGIASVQTNITHAKESHSTPQNIKLVGTYDTSQVDSKTMKQFKEIEKEDNNE
 HITKHGNKVVEDKLPNPNKTSSYSADGSAENNTKVINFSDFVGNMMDGKDDGKISDGIT
 FYSKGSYNGQHDGQVKKGTHVHCNRFNGTKSDHRYWSKKHPRAYVDFYKSDCWYHAKAY
 KCSSLGKMTKCDGLNSIYRKGVKDCSSWKGKPKHKNNWPKTAWYRN*

Sequence 957

Contig_0535_pos_7777_8139,

putative peptide of unknown function

5 atgaataaaaatcttaaaaatattaataacttctattattgttatcattattaccttaaca
gtttggacttttagtggtattacttatcagaaacacaagagtgaagaaatcatcaatcac
gttatagaacgtaagggttgggataaaaaataaaaaatgaaaaatgagttttaattatt
ataatgggatatgctgaaaaagatatgttttttaagatcaaccatatagtgaagtatgag
tataacgtgacaccagcaccatggacagatgataaagaatataaggtgtgggggaaaca
10 gatttacaaaagaaagactcctattataaataatcttttagaatcagaaccttacagaaaa
taa

Sequence 958

MNKILKILITSIIVIIITLVWTFVITYQKHKSEKIINHVIERKGWDKKIKNEKMSFNI

IMGYAEKDIFKDPYSEYEYNVTPAPWTDDKEYKVGWGETDLQKKDSYYKYLLESEPYRK

15 *

Sequence 959

Contig_0535_pos_8145_0,

putative peptide of unknown function

20 atgtataacataataaaaaatagagatcattcatggacttctagtaagattcaaggtaga
aatcacagatgggtgattagaatggtcaccagatcataaatcacttatttataaatatgat
gcaacattaggtagacaaaataaataactaatgacgtgttaactttacttcaagcaacagct
aaaaactcaaatttacgttcaaataatcaatagtaatgaaaaacagtttagcagaacgagg
tctaattgggtatttctaattctataattagagatgatggcgagaaatcttatttacttaac
25 tcaaatcctatttcaagtattagacttagtagaaccagataatgggttacggtggacgtcaa
gtcagtcatttctaacttatataatgaaaaaaattcttctatcgtaaatgggtcaagtt
ccagaagctaattggggcatccgcttttaatatgtataaagttgttaaagctaattgcggca
aataatgggtattatgggtgttatctataaaggcacaattatacttagcaccatacagtcga
aaagggttacattgaaaaattaggccaaaatttaagcaataccaataacgtgattaatgtt
30 tattttgtgccttctgataaagtaaatccttagtataactgtaggttaattacgaccatcat
acggtatatttctgtgaaacatttataaatactatcaatgtaaatgataattatggatta
aataacagtagcttctacaagttagtgacgtgcaattactatgaccagaaaacaacaggtta
gtaggtcagggtcctaattgttactaatagcacaataaaaattgtaaaagttaaagccaca
gataaaagtggaaatgaaagtattgtttctttcacagtaaatataaaaccattaaacgag
35 aaatatagaataacaacttcatcaagtaatacaaacaccagtgagaattagtaatttcaa
aacaatgctaactttcaattgaagatcaaaatagagtaaaatcttcaactcagcatgact
aaaatttttaggtacaagaaattatgtcaatgagtcaataatgacgttcgtagtcaggtt
gtaagtaaaagtataatagaagtgggaacaatgctacagttaatgttacaactacattttct
gatggtacaactaatacaataaccgttccagttaaacatgtgttattagaagttgtacct
40 actactagaacaacagtaagaggacaacaatttccaaccggcaaaaggaacttccccaat
gatttcttttagtttaagaacgggaggtccagttgatgcgagaatagtttgggttaataat
cagggacccgatataaatagtaatacaatttggtagagatttaacattacacgctgaaata
ttctttga

45 Sequence 960

MYNIIKNRDHSWTSSKIQGRNTDGGLEWSPDHKSIIYKYDATLGRQINTNDVLTLLQATA
KNSNLRNSINSNEKQLAERGSNGYSKSIIRDGEKSYLLNSNPIQVLDLVEPDNGYGGGRQ
VSHSNVIYNEKNSSIVNGQVPEANGASAFNIDKVVKANAANNGIMGVIYKAQLYLAPYSP
50 KGYIEKLGQNLNNTNNVINVFVPSDKVNPSITVGNVDHHTVYSGETFKNNTINVNDNYGL
NTVASTSDSAITMTRNNNELVGQAPNVNNTNKIVKVATDKSGNESIVSFTVNIKPLNE
KYRITTSSSNQTPVRIISNIQNNANLSIEDQNRVKSSLSMTKILGTRNYVNESNNDVRSQV
VSKVNRSGNNATVNVTTTFSDGTTNTITVPVKHVLLEVVPTRTRTVRGQQFPTGKGTSFN
DFFSLRTGGPVDARIVVWVNNQGPDIINSNQIGRDLTLHAEIFFX

55 Sequence 961

Contig_0535_pos_5483_5088,

putative peptide of unknown function

atgcaaagaaaatacaaaattataggtattatttttatcgttcttcttattgttttaaca
ttaatttttagtatagtgcatcattatgctaattgttcaaaaacatgaagaagctaaacta

agacaaaaagttcaccatattttcaaacaaaaaggttggaagataaagttaaagaagaa
aaaaatatatttacgttcaatactgggataatgatttacaagtcacttttaagatgag
ccttataatacgtatacactctattgatgaaaacaataaagtatatggacatgctgtt
ttgaaagatgaatatgataaagattttgatagtaaaaaaaagtacaaagaatatttaaga
5 aaaatgcattttgaagaaaaatgatctgaaataa

Sequence 962

MQRKYKIIGIIFIVLLIVLTLIFSIVHHYANVQKHEEAKLRQKVHHIFKQKGWEDKVKEE
KNIFTFNTGDNDLQVTFKDEPYNTYTYSIDENNKVYGHAVLKDEYDKDFDSKKKYKEYLR
10 KMHFEKYDLK*

Sequence 963

Contig_0535_pos_4528_4172,
putative peptide of unknown function
15 atgataaagtactttcagttttattgcaagtttattagttttaaccggattatctttt
tcttcaattaatcaaggtaatactgcgagtgctaaacaaaaatataaaacaactataact
taciaaaggtcaaaaatatgtatatgttggtcattataaacaccatttttctaaaaaagta
gttaaattttctaaaggagttcattcaggaaacaaattagtatctgctgctagtaaaactt
agtaaaaatgggttatgttaaagcatcctcggtattttataaagcctttgattttgattg
20 aaaaatgaattaaaaggtagctatttttacacagcagctaaaaaggtacaggcgtaa

Sequence 964

MIKLLSVFIASLLVLTGLSFSSINQGNTASAKTKYKTTITYKGQKYVYVGHYKHHFSKKV
VKFSKGVHSGNKLVSAAASKLSKNGYVKASSAIYKAFDFGLKNELKGSYFYTAAKKVQA*
25

Sequence 965

Contig_0535_pos_3277_2618,
putative peptide of unknown function
30 atgtcacaaatcgacacaccaatcaactgatcataaaagacaatctcaagataagaaccaa
acacatgacctatgcaaaagtttggttatattttatgtattattggatgatttttgggatt
ggttggtattttggacaatatcttccatgtcatggagaaaaccattgtctctaggattg
ttaattctgatattagcaacgttggtttattaaacgtgcgcgaaaatatggattagtgatc
tctcatatatacgctattattgttaggtttgctttcgatgctctttttacgacttactta
caaaatttaggagcagaagtattctataaaaaatattattcttgctattgggtgcatttata
35 gcttttggtattataggttactttttaataaaagacgcttcgagtatgggaaaaatatttg
tttggtgacattaattgcgctaattatagctgggattatagggatatttattaataatcca
atttttcatactgtcattacaatagtgagcttattattgtttctcctttatactttatac
gattttaatagaatgaaaagaggtcaattctcaccaagagaaatgggatttaatttggtt
40 atcaacttattaatatcatcgaagatatacttagtttagcaaatcgctttaaaaactaa

Sequence 966

MSQSTHQSTDHQRQSDKNQTHAYAKVWLYFMYWMI FGIGCYFGQYLPMSWRKPLSLGL
LILILATLFIKRARKYGLVISHIYAIIVGLLSYALFTTYLQNLGAEVFYKNIILAIGAFI
45 AFGIIGYFLIKDASSMGKYLEVTLIALIIAGIIGIFINNPIFHTVITIVSLLLFLLYTLY
DFNRMKRQGFSPREMGFNLFINLLNIIEDILSLANRFKN*

Sequence 967

Contig_0535_pos_2355_181,
putative peptide of unknown function
50 atggcaaattcttgtttgcataactttcaaaaaaagaatatacggcaacacgatgtcaa
gacggcattttatttatttggcctatcgaaggagtgatgcactttcaacaatttatgaaa
gaaaggatactctcagatgagttatatatttgtaataataggatgtgttttagtatcagt
gacaatggcatcacactagaagtatatatttctagtgattggtttacagagttaggctat
55 tctttttttaattaccattatatatttcggatttaatacaatctaagaaagaaattaaagaa
ctagttgctcaacttacgttgaatttttttagataatgatgtggataaagagcaagatatt
atcaataaaaattgttcatattcttgctaattgaggtcattattgacaaaaaaattgctgaa
gaccaatatatgtatgattattatggtgagttaaaggatgaattgaattatataataat
cacatcgaagaaagacttactctaaaagatatttccaataaattatatgtttctaagtcc

aacctttctacgcagtttcatttggtattaggtatgggatttaaaaaatatatagataca
 cttaaaattagtaaatctatagagatgctacttacgacaactaaaacgataagccaaatt
 agtgaaacgtaggatttagtaatgtatctacatattctagacaattcaaaaattattta
 agtgtaacgccaatgcatatcgtgcaatgaaaaaatatgataagtacaatggatgttct
 5 gatgatgatgtttcagaacacttaaaatcatgtgtacaatcattaatatgttctaaaatg
 ccaacgaatgagttagataattatgatgaaattgttattgaccaatatccaatttcta
 gtttcaacgttttattctgtcgttcaaattaattcaattgatgaaattaaaatgttgttt
 ttgcaaggatttcataaaaaaatagggtatgaagggtcaaataatttttttgtattatg
 cctaacttatgccaatataagaatttgttctctcaagaagagatgaatgatattcattaaa
 10 attattattgaatatcgcttgcacgtcgcatttagcatcgataaaatcgagcaaatatat
 gaacttaatacaatttttacatatcaatatgaaaacttaaaaattatgaataaatgttct
 gtttcagattacaattgtgcaatttatatttaatttgaacgaaaaagatttcgagaaatt
 tatcgaatatcttgaagatacaaaacatcgaattggaatataaaaatagggttagatatt
 agttgcatgtttaatgacactgcacaatttaaatcattagcttcgaaaataaagcggtta
 15 aaatttgactatctttatattgataatgctagattaaagtcaccctatttacttgataat
 gaggaagggtttattactcaaaaaatattctacaatttaagcatttaatagatgatttaaa
 cagtttgatttttagtagtgaaaatttaatttttctaaatttatataatcatcaattactg
 aataataatgaaattgatttaagtaataagcgctccattactatttaaacgatttcaaaa
 ctaaaaaaacatttttaaggctacggattaaatgtattttcaaatcctaaagtccttaac
 20 gctgtacattttatttgatgagaatgggtttaaaacaacgtttgactgatttttaatcat
 ttgagttggatgactaatcaaaaccaaattgaacaacgattctataatattattgaaaat
 gctgatcaatattatcttacttatatgattggcgtgtgattgaaagtgaatctaagag
 agcgactttaagacggttgatatatggattaactttgaagatgaagcgtaatagatgaa
 tataatttggtgtgattgctaaagttgatgatgaagggtggcaatattaatcatatgatttct
 25 caaaacttacgtcacaaatatgtttggtctacaccggttcttgatgagagttgaggagaac
 tttagaccatacatgcatttatggaacatgactttaaaaaagggccattgaaaatcaga
 atgaaatataatgcagtatatgtagttgaaatatataaaaaagataaaaataaaaag
 cgtagcacaaacttaa

30 Sequence 968
 MANSCLHILSKKEYTATRCQDGIILFWPIEGSMHFQQFMKERILSDELYIVNNMDVFSIS
 DNGITLEVYISSDWFTELGYSTFFNYHYISDLIQSKKEIKELVAQLTLNFLDNDVDKEQDI
 INKIVHILANEAIIDKKIAEDQYMYDYYGELKDELNYIYNHIEERLTLKDINSKLYVSKS
 NLSTQFHLLGGMFKKYIDTLKISKSIEMLLTTTKTISQISETLGFSNVSTYSRQFKNYL
 35 SVTPNAYRAMKKYDKYNGCSDDDVSEHLKSCVQSLICKMPTNELDNYDEIVIDQYPISN
 VSTFYSVVQINSIDEIKMLFLQGIHKKIGYEGSNIIFCIMPNLCOYKNLFSQEEMNDI IK
 IIEYRLHVAFSIDKIEQIYELNQLFTYQYENLKIMNKCSVSDYNVQFIFNLNEKSIREI
 YRNILKIQNIELEYKIGLDISCMFNDTAQFKSLASQIKRLKFDYLYIDNARLKSPYLLDN
 EEEGLLKNILQFKHLIDDLKQFDFSSENLI FLNLYNHQLLNNNEIDLNSAPLLFKTISK
 40 LKKHFKGYGLNVFSPKVFENAVHLFDENGFKTTFGLIFNHLNMTNQNQIEQRFYNIEN
 ADQYYLYLYDWRVIESESNESEDFKVDIWINFEDEALIDEYICVIAKVDEGGNINHMIS
 QNLRHKYVWSTPFLMRVEENFRPYMHIMEHDFKKGPLKIRMKYNAVYVVEIYKKDKINKR
 RSTT*

45 Sequence 969
 Contig_0536_pos_1715_2401,
 is similar to (with p-value 7.0e-31)
 >sp:sp|P39787|DNAD_BACSU DNA REPLICATION PROTEIN DNAD. >gp:g
 p|L47709|BACYPIA_25 Bacillus subtilis (clone YAC15-6B) ypiAB
 50 F genes, qcrABC genes, ypjABCDEFGH genes, birA gene, panBCD
 genes, dinG gene, ypmB gene, aspB gene, asnS gene, dnaD gen
 e, nth gene and ypoC gene, complete cds's. NID: g1146223. >g
 p:gp|U11289|BSU11289_2 Bacillus subtilis 168 asparaginyl-tRN
 A synthetase (asnS) and endonuclease III (jooB) genes, parti
 55 al cds and DnaD protein (dnaD) and (jooC) genes, complete c
 ds. NID: g533096. >gp:gp|Z99115|BSUB0012_175 Bacillus subtil
 is complete genome (section 12 of 21): from 2195541 to 24092
 20. NID: g2634478.
 atggatctaattcaattaaaaacaagacctgttggttataagacgagaattgtttgatcat

tattcagagttaggtttggatgaacaagatttagttattttgataaaacttttatatgca
 tctgaaacttctaataagcaaccttctattgaatttcttcaaaaaggatcaactatggaa
 cctcgtcaaattacttccgtaatacaaaaacttaattcaagagaattattagaactcaat
 gtttagtaagacgaagaaggtaaaattcactgaatacatgaatttggatcccttctatcac
 5 aaattaaatcaattattaaaacatcaatacttaaaacatgaggaacaagataaaaaagag
 cagtttaagcaattgtttcagatagttgagcaatcgttcggcagaccactatcgccgtat
 gaaattgaaacattaaatcagtggttggatgctgatcaccatgacttatcagttatacaa
 gccgctcttgatgaggcacttagccaaaataaacttagttttaaatatattgatcgtatt
 ttattaaattggaaaaagaataatgtgaaaacagttgacgattcaaagaaaataagagaa
 10 cagtttaacaaacaaaatgaaacatgttgtaaaaaagggtgcctaaatttgactggttg
 aatggagagaatcctaataagataagtaa

Sequence 970

MDLIQLKTRPVVIRRELFHDHYSELGLDEQDLVILIKLLYASETSNKQPSIEFLQKGSTME
 15 PRQITSVIQNLIQRELLELNVSKEEGKFTEYMNLDPFYHKLNLQLLKHQYLKHEEQDKKE
 QFKQLFQIVEQSFGRLSPYEIETLNQWIDVDHDLVSIQAALDEALSQNKLSEFKYIDRI
 LLNWKKNVKTVDSSKKIREQFNKPKMKHVVKVPKFDWLNGENPNDK*

Sequence 971

20 Contig_0536_pos_2418_3050,
 is similar to (with p-value 8.0e-76)
 >sp:sp|P39788|END3_BACSU PROBABLE ENDONUCLEASE III (EC 4.2.9
 9.18) (DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE). >gp:gp|L4
 7709|BACYPIA_26 Bacillus subtilis (clone YAC15-6B) ypiABF ge
 25 nes, qcrABC genes, ypjABCDEFGH genes, birA gene, panBCD ge
 nes, dinG gene, ypmB gene, aspB gene, asnS gene, dnaD gene, n
 th gene and ypoC gene, complete cds's. NID: g1146223. >gp:gp
 |U11289|BSU11289_3 Bacillus subtilis 168 asparaginyl-tRNA sy
 nthetase (asnS) and endonuclease III (jooB) genes, partial c
 30 ds and DnaD protein (dnaD) and (jooC) genes, complete cds.
 NID: g533096. >gp:gp|Z99115|BSUB0012_174 Bacillus subtilis c
 omplete genome (section 12 of 21): from 2195541 to 2409220.
 NID: g2634478.
 atgattgacgttatagcagatatgtttcctaatagcagaatgcgaattaaacatagaaat
 35 gcattcgatcttacaatagctgtattattatcagcacagtgtactgataatctagtcaat
 cgtgtcactcaatcattatattagaaaaatcgaacacctgaagattattaaatgtgagt
 gatgaagaattacaaaatgatatacgctctattggattatcgcaataaagccaaaaat
 ataaaaaattatgcccactctttaattgaacaatttaattgggtcaaatccacaaacacat
 aaagaattagagagctctagctggagtggggcgtaaaacagcaaatgttgtaatgagtgtc
 40 gcatttggagaaccttcttttagctgtcgatactcatgttgagagagtttctaaacgtttg
 ggaattaatcgttggaagatagtgtaagacaagtagaagatcgattatgtgatattatc
 ccaagagatagatggaataaaagccatcatcaattaatatttttgggagatatcattgt
 cttgctagaaaacctaaatgtgagatatgtccgctgttaaatgattgtagagaaggacaa
 aaacgacataaagcaagataaaggagcgctga

Sequence 972

MIDVIADMFPNAECELNHRNAFDLTIIVLLSAQCTONLVNRVTQSLFRKYRTPEDYLNVS
 DEELQNDIRSIGLYRNKAKNIKKLCHSLIEQFNGQIPQTHKELESAGVGRKTANVVMVS
 AFGPSLAVDTHVERVSKRLGINRWKDSVRQVEDRLCDIIPDRWNKSHHQLIFFGRYHC
 50 LARKPKCEICPLLNDCREGQKRHKAKIKEA*

Sequence 973

Contig_0536_pos_3055_3387,
 putative peptide of unknown function
 55 atgattgaaaaacaggatttcaatcatatagaggaccaacttgatcaactagcaagtaat
 aaacaactcaaaaacaccagaagctagggaacttttagatagttatttcgatttaattatt
 aattatttttaacaaaataaataacatagatgaaattcatttttaataactcgatacatat
 ccagtagttccaatgaattttgatgaacgctatcattatatggttgacgtaaacaccat
 tttatgggctatcgtaaatgaaaacattgaaatcagaattaataaaaatgaatgcatct

tatctaattagaaagcaacgtcaacaaaaataa

Sequence 974

5 MIEKQDFNHIEDQLDQLASNKQLKTPEARELLDSYFDLIINYFKQINNIDEIHFNQLDITY
PVVPMNFDERYHYMVARKHHFMGYRQMKTLKSELIKMNASYLIRKQRQK*

Sequence 975

Contig_0540_pos_1272_2228,
is similar to (with p-value 6.0e-27)
10 >gp:gp|AF076683|AF076683_2 Staphylococcus aureus oligopeptid
e transporter putative substrate binding domain (opp-1A), ol
igopeptide transporter putative membrane permease domain (op
p-1B), oligopeptide transporter putative membrane permease d
omain (opp-1C), oligopeptide transporter putative ATPase dom
15 ain (opp-1D), and oligopeptide transporter putative ATPase d
omain (opp-1F) genes, complete cds; and unknown gene. NID: g
3800817.
atgctcaaacgtacaattaaattcatactttattttaatcgtaagttcggtttattatcttc
attttagttgagaagacatctggaatccagcgattctgtatctacaacgtcatggttat
20 acgtcgattacgcaagacaatattgaagcggcacaacatcaacttggttaggacaacat
gtgttactaagatatatcgattgggttgacatgcactcacgggcaacttaggatacggc
tttagtacgaacgaagcagttaccgctatgataatggaagccatcggtccgacgcttggt
ctaactcattgtctctagttgtatcatgttgccatttggtatattgttggttacttcgtt
ggacgcgctccgcatacacggttacgctaattggaattcggtgattcgcccaagtgatgacc
25 tcaatgccagaatactggtagctattttattcatttattatttaggcgtacgttggtgaa
ttgttaccatttgtaggtagtgattcatggcaacactttgtgctgccaatcttcacaatt
gttggttagaagggtgtcatatcttattgatgacagcacatctgattacacaaacgtta
gatcaagatgcgtatcaactggcgagtttaagacatttttcgttaaaagcggtatcatc
gtacaaattaaagagatatattgcaccactaatgacgatttcaattaacagtatcattcat
30 ttaattggaaaagccgtaataactagaagtcattcttcagcatgtctggtataggttaaattg
ttgattaatgctattaaccaacgagattatccactgattcagggcattgtcatctttatc
attgtctttattatgctaattgaattatttaggcgatgtgattattttgaagaatgaacct
agacttcgacgacgtcataccagcagtcaggcaatgagaaaagaggtacgatgtga

35 Sequence 976

MLKRTIKFILYLIVSSFIIFILVEKTSGNPAILYLQRHGYTSITQDNIEAAQHQLGLGQH
VLLRYIDWVGHALTGNLGYGFSTNEAVTAMIMEAIVPTLVLIIVSSCIMLPFGYIVGYFV
GTRPHTRYANGIRGFAQVMTSMPEYWLAILFIYYLGVWRWQLLPFVGSDSWQHFVLPIFTI
VVIEGCHILLMTAHLITQTLDDQAYQLAQLRHFSKARIIVQIKEIFAPLMTISINSIIH
40 LIGKAVILEVIFMSGIGKLLINAINQRDYP LIQGIVIFIIVFIMLMNYLGDVILKNEP
RLRRRHTQSGNEKRGTM*

Sequence 977

Contig_0540_pos_2276_2995,
45 is similar to (with p-value 5.0e-32)
>gp:gp|U64514|BFU64514_3 Bacillus firmus dppABC operon, dipe
ptide transporter protein dppA gene, partial cds, and dipept
ide transporter proteins dppB and dppC genes, complete cds.
NID: gl813494.
50 atggttgatttaattacgtatggtttaatgcaagacacgcaacatttgaaacccacttgag
tcacctaattggacaacattggttggtaccgatcaattaggcagagacttcttagtaaga
ctgattgtcgttagtcttgatcattgagtttaacaggcatagtgattctattaagcgtt
tgtatgggacttatctttggcttaattgcaggcatagaaagacgatggttagatcaaact
atcatgtttgttgccgatatgttgctggctattccgctcatttattatcgcatagtcac
55 ttaagtttagtaagtaactccatgataggtttgatacttgctttaacgattggatggata
ggacgttatctacgttacttcagaaatttaacgcgagatatcaaaaacgtccatttggt
caatatgcacgattgagtggaactcaacattcaaaacgcagtaacacatgtgattcca
catttattaagtagtatattcgctttggttaacggctgactttggcaaaatgatgctcagc
atatctggacttgcttttctaggactaggtattaaaccgcgcacgcctgagttaggaaca

attctttttgatgggaaaagttatttcaacggcgaccgtggctcttcttcttccttgggt
gtattgttaggaggtttcgccttattatgtcaaattatcaacaaaaataacgcagtaa

5 Sequence 978

MVVLITYGLMQDTQHLNPLESPPNGQHWLGTDLGRDFLVRLIVGSLVTLSTGIVILLSV
CMGLIFGLIAGIERRWLDQIIMFVADMLLAIPSFIIALVILSLVSNMIGLILALTIGWI
GRYLRYFRNLTRDIQKRPFVQYARLSGNSTFKTTVTHVIPHLSSIFALVTADFGKMMLS
ISGLAFLGLGIKPTPELGTILFDGKSYFNGAPWLFFFPGVLLGGFALLCQIINKKITQ*

10

Sequence 979

Contig_0540_pos_3148_3729,

is similar to (with p-value 2.0e-19)

15 >sp:sp|P45095|DPPD_HAEIN DIPEPTIDE TRANSPORT ATP-BINDING PRO
TEIN DPPD. >pir:pir|F64188|F64188 dipeptide transport ATP-bi
nding protein (dppD) homolog - Haemophilus influenzae (strai
n Rd KW20) >gp:gp|U17295|HIU17295_3 Haemophilus influenzae d
ppB, dppC, dppD, dppF, isn, artP, artI/J, artQ, and artM gen
20 es, complete cds, and opa gene, partial cds. NID: g972894. >
gp:gp|U32798|U32798_3 Haemophilus influenzae Rd section 113
of 163 of the complete genome. NID: g1574110.

atgaaacaatcacaattatgttatcaaggagatattgacatcgatttaactcaaacagat
gcagtgtttcaagatgttcaaaagtaatatgtttcaaaatataacattagctaagcatttc
25 caatacatattatgaagccaatcgacacatctcactaaacagcgtatttaaggaagatgtc
ttacagatgatgcaattacttggtttaagacaaggggaacaattgcttgagcgttatccc
ttcgaacttagtggaaggtatggcacaacgtgtcgcctttataatgtcattaattagacgt
ccgaactacttatttttagatgaaccaacgagtgacattgatcaagaaaatattaaaaag
tttatgcattaccttcttagggcacaggagcgctaccaaatagaccattgtttttatcaca
30 catgatattaacttagtgaaaagattgtgccacacatattagattatgcagcaaggtaaa
ttgatagaaaatggtagggccgcgtcgatcttaactaagccgacacataattacacgaaa
aattaattacgattgcacatcgagacaaccttatgtcttaa

Sequence 980

35 MKQSQLCYQGDDIDLTQTDAVFQDVQSNMFQNI TLAKHFQYIYEANRTHLTKQRIKEDV
LQMMQLLGLRQGEQLLERYPFELSGGMAQRVAFIMSLIRPNYLFLEPTSALDQENIKK
FMHYLLRAQERYQMTIVFITHDINLVKDCATHISIMQQGKLIENGEAASILTKPTHNYTK
KLITIAHRRQPYA*

40 Sequence 981

Contig_0540_pos_5018_5803,

putative peptide of unknown function

atgcaacattcaagcaaaataatagtatattgttaagtttcttaattttaacgatttttatt
ggaggatgtggttttataaataaagaagatagcaaaagaacggaatcaaacaaaacttt
45 aataaaaatgttagacgtgtatccaactaaaaatctagaagacttttatgataaagagggc
tatcgtgatgaagagtttgataaagatgacaaaggaacatggattattaggtctgaaatg
acaaaacagccaaaaggtaaaattatgacctcaagaggtatggttctctatatcaatcgc
aacactagaacagccaaaagggtatttttaataagataaagataaaagatgatagtaattggt
agaccgatagagaatgaaaagaataaccctgtaaaaatgaaccataataagatctttcca
50 acaaagccaatatctgatgataagttaaaaaaagaattgaaaacttcaaattttttgtg
caatatggaaatttttaaaaacttaaaaggattataaaaacggggatattttatacaatcct
aatgttccctagttattctgcgaataatcaattgagtaataatgaatataacgtacaacaa
ttaagaaaaagatatgacatcccaactaaaaaagcacctaaactattgttaaaaggggat
ggcgacttaaaaggatcatccgtaggtcatagagacctagaatttacctttgtagagaat
55 aagaaagaaaacatcttttttacggatagtttaatttttaaccgactgagcgtgatgaa
tcatga

Sequence 982

MQHSSKIIIVFSFLILTIFIGGCGFINKEDSKETEIKQNFNKMULDVYPTKNLEDFYDKEG

YRDEEFDDKDDKGTWIIIRSEMTKQPKGKIMTSRGMVLYINRNRTRAKGYFLIDKIKDDSN
RPIENEKKYPVKMHNHNIPTKPIISDDKLKKEIENFKFFVQYGNFKNLKDYKNGDILYNP
NVPSYSAKYQLSNNEYNVQQLRKRYDIPTKKAPKLLLKGDDLKGSSVGHRLLEFTFVEN
KKENIFFTDSINFKPTERDES*

5

Sequence 983

Contig_0540_pos_6075_7043,

is similar to (with p-value 2.0e-20)

>gp:gp|AJ222587|BS16829KB_25 Bacillus subtilis 29kB DNA frag
ment from ykwC gene to cse15 gene. NID: g2632216. >gp:gp|Z99
111|BSUB0008_93 Bacillus subtilis complete genome (section 8
of 21): from 1394791 to 1603020. NID: g2633699.

atgacttactgctcattatctataaagatttatacattgaaaatattaattgttttaaca
ataggaggtcatgttgttattatgagtcagtttaaggacacattatataaactatttgag
ccaatgatgaaaatagagttctatcaaaatcttttggttaattcttttaattatacttgct
tatactcttgatgggtatgattgtaattgcgatatcaagaaagttagttactaaatttttc
aacgttaatgaaaagaaaagaaccgtcataaaattaagagaagtgaacactatccaca
ttgattcaaaatttaataagttatgtcgtatggtttattgtccttacgtcaataactttca
cgtttcgggtattagtgatcagcaatttttagcaggagctggagttgttggtgttgccgtt
ggtttcggagcacaacaattgttaaaagacattattactggtttctttatcatatttgaa
ggacagtttgatgtgagtgattatgttcaaattaatgcatctggggtaacaattgctgaa
ggtagcgttaaaacgattggtttaagatcaacgcgtatacaatcagatactggagaaatt
tatacattacctaattggtatgattagtgaaatagttaattattctgctacagatgtttca
cctattgtgatgataccgatttctccaaatgagaattataaagtgatagaagagaaatta
ttaacattttttacctacattaaagaataaataatgatcatatttgtatccgcaccagattta
cttggttttagatagtggttgatggcaatgaaatgggtgattaaacttttagcacatgttaag
cctggaatgcattttccaggacaacgtttacttctgtaagagggtcatataacttttagt
gaagaaggcattcatattccgaaaccaacacttgtaaaacttgataaagaattgaataaa
aaagaatag

30

Sequence 984

MTYCSLSIKIYTLKILIVLTIGGHVIMSQFKDTLYKLFEPMMKIEFYQNLLVNLLIILA
YILMGMIIVIAISRKLVTKFNNVNEKKNRHKIKRSETLSTLIQNLSYVWFIVLTSILS
RFGISVSAILAGAGVVGAVGFGAQTIVKDIITGFFIIFEGQFDVSDYVQINASGVITAE
GTVKTIGLRSTRIQSDTGEIYTLPLNGMISEIVNYSATDVSPIVMIPISPENYKVIEEKL
LTLFPTLKNKYDIFVSAPDLLGLDSVDGNEMVIKLLAHVKPGMHFPGQRLLRKEVIQYFS
EEGIHIPKPTLVKLDKELNKKE*

35

Sequence 985

Contig_0540_pos_685_182,

is similar to (with p-value 4.0e-46)

>sp:sp|P54452|YQEG_BACSU HYPOTHETICAL 20.1 KD PROTEIN IN NUC
B-AROD INTERGENIC REGION. >gp:gp|D84432|BACJH642_91 Bacillus
subtilis DNA, 283 Kb region containing skin element. NID: g
2627063. >gp:gp|Z99117|BSUB0014_48 Bacillus subtilis complet
e genome (section 14 of 21): from 2599451 to 2812870. NID: g
2634966.

atgccaaatgcatatgtgaaatcaatatttgaaattgatatagaaaaacttgccgatagt
ggtgttaaagggtatcataactgatttagataatacacttggttggtgggatgttaaagaa
cctactaagggtgttaaatcatggtttgctaaggctaaagatttaggaataactgtcaca
attgtgtcaaaataataaaaagtcgagtatcaagtttctcaagtaatttaggtgtagat
tataatattcaaagcacgtaaacgatgggaaagcctttaagatggctattaaaaaaatg
aaaattcaaccgagagaaaccggtgttgtaggagatcaaatgcttactgatgtgttggt
ggcaattgtaattggtttatatacaattatggttagtacctgttaaacggactgatggatta
attcaaaagtttaactgattaattgaaagacgattattaaatcatttttagaaaaaaaggt
tatattaaatgggaggaaaattga

55

Sequence 986

MPNAYVKSIFEIDIEKLADSGVKGIITDLDNTLVGWDVKEPTKGVKSWFKAADLKITVT

IVSNNNKSrvSSFSNLGVdYIFKARKPMGKAFKMAIKMKIQPRETVVVGdQMLTDVFG
GNCNGLYTImVVPVKRTDGLITKFNRLIERLLNHFRKKGYIKWEEN*

Sequence 987

5 Contig_0541_pos_1165_1485,
putative peptide of unknown function
atgacaggaagaacgaataatacgaataaccacgccaagttgaactggcagtagctcga
tcacgttcaaggaaaatgatgacaaatgccaaatactacgttagttataaaacccaactact
aacaatatagttaaaatagtacccaaaaatccaaaattcattactaaataactcctattca
10 gtctttcaccaagaaaatttagccgccgatttactgaactatgcttccgttaccaatgtt
aaaatagcattaaagtccctcgctcattatgcggaaatgcaggttcttcgatttggttaata
gtaatgtttaataatttgtaa

Sequence 988

15 MTGRTNNTNNHAQVELAVRRSRSRKMMTNANTTLVINPTNINIVKIVPKNPKFITKYSYS
VFHQENLAADLLNYASVTNVKIALKSSSLCGNARSSICLIVMFNNL*

Sequence 989

20 Contig_0541_pos_3076_3672,
putative peptide of unknown function
atgtggagggtatgcgcattggaagaattgggttaaaattgactgtttcatcagtcctatatt
attagtttagtttaaacactttttatctcaagttagtctattaaatgagaataaaacaaat
caaatagaacatgcatcaactatgaaagaaaagtctaatataaataatgtaaaaaaact
aaaaataaaaaatattgaaaaatcaacgcagacagacaaaacaaaactctgtgaacttaag
25 caaaacacaaaagatcaaaataataacgcgaatgatgaagcagcttctccaactagcgaa
caaatgcagctatagcacaagcaaagtcattatgcaatacattacctatctctaagaaa
agtttatacaacaattaacttcggaatacggagagaaatatccggcagacatagcacag
tatgctgttgaccatatcagtgtagattataaaatgaatgcaactgagattagcaaaaagt
tacgtaaaaaatataaacatttctaatacgaagcgttatatgatcaactcgtttcagaaaat
30 ggagaaggatttactcctgaagaagcacaatatgcaatgaatcatttagataggttaa

Sequence 990

MWRYAHWKNWLKLTVSSVYIIISLVLTLLFQVSLLENKTNQIEHASTMKEKSNINNVKTT
KNKNMEKSTQTDKQNSVNLKQNTKDQNNNANDEAASPTSEQNAIAQAKSYANTLPISKK
35 SLYKQLTSEYGEKYPADIAQYAVDHISVDYKMNALRLAKSYVKNINISNQALYDQLVSEN
GEGFTPEEAQYAMNHLDR*

Sequence 991

40 Contig_0541_pos_10591_9902,
is similar to (with p-value 4.0e-42)
>sp:sp|P54471|YQFN_BACSU HYPOTHETICAL 23.7 KD PROTEIN IN CCC
A-SODA INTERGENIC REGION. >gp:gp|D84432|BACJH642_139 Bacillu
s subtilis DNA, 283 Kb region containing skin element. NID:
g2627063. >gp:gp|Z99116|BSUB0013_228 Bacillus subtilis compl
ete genome (section 13 of 21): from 2395261 to 2613730. NID:
45 g2634723.
atgattaaccttaaccaaagattatcaattgtatgctcatttattaaaagaggaacattg
gctgatattggctcagaccacgcatactctacatatatgcaattcaaaacgacttatgc
acaaaagcaatagcgggagaagtgattcaaggaccttataaggctgctaaaagaaatatt
50 gcaaatatgaattaaatcaacagggttgatgtacgtctaggcgatggtctaagcggtata
aacacagaagaccaaattgataataaactgtttgtggtatgggagggccattaattgca
aaaatattaaacgatggaaaagataaattagtttaaccatccaagactcatactacaaagc
aacatacaaaactcaagcattaaagacaaactcttaataaactttcatatgaaatcggtgat
gaaagaatcattgaggaaaagggtcacatatatgaaatcggtgtagctgagtttaataat
aacttagtttaaatataattacaagaaaaattcggaccatttttacttagagaatgt
55 aataacatttttcaaaaaaatggcaaagagagttagaagcactgcgtgatataaaatcc
caattgaattcaacatcacatcatgagagactaaaagaaatagaagatgaaattaactta
atacaagagggtgtaattaatgaaaattag

Sequence 992

MINLNQRLSIVCSFIKRGTLADIGSDHAYLPIYAIQNDLCTKAIAGEV IQGPYKAAKRNI
ANYELNQQVDVRLGDGLSVINTEQIDNITVCGMGGPLIAKILNDGKDKLVNHPRLILQS
NIQTQALRQTLNKLSEIVDERIIIEKGHIYEIVVAEFNNNLVKLNILQEKFGPFLREC
5 NNIFQKKWQRELEALRDIKSQLNSTSHHERLKEIEDEINLIQEV LINEN*

Sequence 993

Contig_0541_pos_9547_8813,
is similar to (with p-value 7.0e-32)
10 >sp:sp|P53434|YRP2_LISMO HYPOTHETICAL 41.4 KD PROTEIN IN RPO
D 3'REGION (ORFA2). >gp:gp|U17284|LMU17284_3 Listeria monocy
togenes major sigma factor (rpoD) gene, partial cds, and dow
nstream orfA1 and orfA2 genes, complete cds. NID: g687597.
15 atgattaatacaaaatagctcatattattacaaagttcaaactttttatacctaaaaattat
attgaagattttcaaagacagtttaaacgaacttggattagctaaagaaggtaattacgaa
tattgtttctttgaaagtgaaggtaaaaggcaatttaaacagtaggtgatgcaagtcct
tatatagggaagtttagatagtagtgcgaatatgttgatgaaataaaacttgagtttatgata
aaagacaatgaattagaaataactaaacgtgctatttttagataatcacccatacgaaca
ccagttttttgattttattaaaatgaacaaagaaagtgaagtatggattagggtatttgga
20 caattaaaccaaactatgacttttagatgaattttctgaatatgccaaaaaacagctcaat
ataccgagcgtacgatatacaggtcaacatgatagtcgaattaaagaaagtagctatcata
ggtggttcaggtataggatttgagtataaagctagccaacttggagcagatgtttttggt
actggtgatattaaacacccatgatgcttttagatgctaaaatccaaaatgtaaatattatta
gacatcaatcattatagtgagtagtatttatgaaagaaggattaaaagaattattagaaaaa
25 tggttatttaaatatgaaaatcaatttccaatatatgcttctgaaatcaacacagatcca
tttaaatataaataa

Sequence 994

30 MINTNSSYYYKVQTFIPKNYIEDFKDSLNLGLAKEGNYEYCFESEGKQFKPVGDASP
YIGKLDSIEYVDEIKLEFMIKDNELEITKRAILDNHPYETPVDFIKMKESEYGLGIIG
QLNQMTLDEFSEYAKKQLNIPSVRYTGQHDSPIKKVAIIIGSGIGFEYKASQLGADV FV
TGDIKHHDALDAKIQNVNLLDINH YSEYVMKEGLKELLEKWLFKYENQFPIYASEINTDP
FRKYK*

35 Sequence 995

Contig_0541_pos_8778_7432,
putative peptide of unknown function
atgtcaaaacatccatttgaacactttaatttagatgagaatttaattgaagctgtttaa
aatctcaatttttgaaaaaccgactgaaatccaaaatagaatcataccgagaattctttaa
40 ggaacaaatttaataaggacaatctcaaactggaactggaagtcacacgcttttcttta
ccattaattcaacttatagaaagtacattcaagagccacaagccatcgtagtagtcca
acacgtgaacttgctcagcaactatcaagttgctatgcatttagttaaattcaaaaaa
ggtataaatgtaaaacttttctattggtggtaccgatttagaaaaagataaacaacgatgt
agccatcaaccacaactcattattggtacaccaacaagaattaatgatttagcacattca
45 ggttatcttcatgcacatttagcgtcatatttaattatagatgaagctgatttaattgatt
gacctcggtctcattgaagatggtgacctattgagcgagatttagatgatgaaaatgct
catctagcggatttttagtgcaacaatacctaaatcattacaaccatttttaataaatat
ttaagtcaaccagaattttgtagaagttgatggcaagctcataataaagaaaatatcgaa
ttttatctaattcctacaaaagggtctgctaaggtagataaaacattggaattgatagat
50 atattgaatccttatctatgtatttttctgtaacagtcgtgaaaatgccgatgaattg
gcagacacttttaataaagaaggaaataaaataggatgattcatggtggtttaacacca
agagaacgtaaacacaacaatgaaaagaataagaaatttagattttcaatttgcattgca
agcgatcttgcttcttagaggaatagatattgaaggcgtgaagtcattgtatttaatttcg
gtacccaatgatatcgatttcttcacacatcgcgtaggtcgaacaggaagaggtaattat
55 aaagggtgtagccattacattatatagtcctgatgaagaaagtaataattactcttattgaa
gacagagggtataaatttgaaaatgtagatattaagaatgggtgaattaaaaccgataaag
gcatacaatatgcgtaaatcaagacagcgcaaagatgaccatttaacaaatgaagttaaa
cacaagtaagaagtaaatcaaacgtaaagttaaacagggtataaaaaagaagtttaaa
caagaagttgaaaaaatgaaacgtcaagaaagaagcagtagtagtaaaaaagcaaaataga

caaaaacgaaaaataataaaggatag

Sequence 996

MSKHPFEHFNLDENLIEAVKNLNFKEKPTEIQNRRIIPRILKGTNLIGQSQTGTGKSHAFLL
 5 PLIQLIESDIQEPQAIIVVAPTRELAQQLYQVAMHLVKFKKGINVKLFIGGTDLEKDKQRC
 SHQPQLIIGTPTRINDLAHSGYLHAHLASYLIIDEADLMIDLGLIEDVDHIAARLDDENA
 HLAVFSATIPKSLQPFNLKYLSQLPEFVEVDGKAHNKENIEFYLIPTKGS AKVDKLTLELID
 ILNPYLCLIFCNSRENADELADTLNKEGIKIGMIHGGLTPRERKQMKRIRNLDFQFVIA
 10 SDLASRGIDIEGVSHVINFDVPNDIDFFTHRVGRTGRGNKGVAITLYSPDEESNITLIE
 DRGYKFENVDIKNGELKPIKAYNMRKSRQRKDDHLTNEVKHKVRSKSRKVKPGYKKKFK
 QEVEKMKRQERKQYSKKQNRQKRKNNKG*

Sequence 997

Contig_0541_pos_5360_4650,
 15 is similar to (with p-value 1.0e-20)
 >gp:gp|271552|SPADCA_3 Streptococcus pneumoniae adRCBA oper
 on. NID: g3758891.
 gtgacattaggtggtatttcccttgggtatttttgccttaccattattcccggttttctca
 gtaataaacccctatgtggttgggtatttcttctgctgttattggagcggtatttaattgaa
 20 aaattaaggacttcgttttctaattatcaagaaattgcaattcctattataatgagcgct
 ggtattgctctaagtgcctatttttatttctctagcagatgggttttaatacaagaaatcgta
 gacctactatttggatcaattagtgacagtaaatattagtgatttaactacaattattatc
 attacaataattgttctcatatttattgttttattttataaagaattgtttattttatca
 tttagcgaagaatatagtaaggctcataggtataccaaagggtgattcaatttttatttata
 25 gtaattgttgctatggaatatctgcatcaatgagagttgttaggtatattatttagtaagc
 gcgttaataactcttctcatagcaatttcaatgagaataactaaaggatttaacaatta
 atagcattaaagtgttatattaggagaattatctgtaattcctaggatttaattatagctttt
 tatatgaatatatcacctgggtggcgctattgttgactattgggtattaatgctcatacta
 acgatgattattcagaaggttaaaaattaagtttaaaaaggagtcggttaa

Sequence 998

VTLGGISFGMFLLTIIIPVFSVINPMWFILFAVIGALLIEKLRTSFSNYQEIAIPIIMSA
 GIALSAIFISLADGFNQEIIVDLLFGSISAVNISDLTTIIITIIIVLIFIVLFYKELFELS
 FDEEYSKVIGIPKWIQFLFIVIVAMVISASMRVVGILLVSALITLPIAISMRITKGFQKL
 35 IALSUILGELSVILGLIIAFYMNISPGGVIVVLLVLMILTMIIQKLKIKFKKGVV*

Sequence 999

Contig_0541_pos_0_1325,
 40 is similar to (with p-value 1.0e-80)
 >gp:gp|U88888|BFU88888_2 Bacillus firmus MecA homolog (mecA)
 and cardiolipin synthase (cls) genes, complete cds. NID: g2
 952026.
 atgaattttggatttttgggtactattttaactatattggttagtagttgggtttataact
 aacgtagtattggcatttgcctatttcttgaacgtgatcgactgactgccagttca
 45 acttgggctggttattcgtattattcgttcttctcgtcattggatttattttgtatcta
 tttttaggacgaacggtttccaagaaaaagatggaaaaataaacggtgatgaattacat
 gcatttgaagatttagttcaagaccaaaccgacagttttgataaacataattatggttat
 atcaatgatcaagtcatttaaacaccgtgatttaatacgtatggttgtaataaacaagat
 gcctttttaacagaaaaataaaaaatcgatttattttacagatggtcataagctttatgaa
 50 aaagtacttgaggatatttacaatgctcaagactatatacatctagagtactataccttt
 gaacttgatggattaggtaaaagaatcttagatgcacttgaaactaaactaaagaagggt
 ttagaagttaaacttttgtatgacgatgttggttctaaaaagggttagattatcaaaattt
 aaacatttcagagcattaggtggagaagttgaagcatttttcccttcgaaagtaccttta
 atcaatttcagaatgaataatcgaaatcatagaaagattatcattatagatggacaaatt
 55 ggttacgttggcggttttaattgctggcgatgattatttaggattaggtaagttagggttac
 tggagagatacacatacacgtgttcaaggtgaatgcacgatgcactacaattaagattt
 attttagactggaattcacagtcgcacgtccacaatttaatttgatcaaaaatatttc
 ctaaaaaaaatggggacaaaggaaacgcggctattcaaatcgcttctagtggacctgca
 tttgatttacatcaaatagaatatggttatacaaaaatgataatgagcgctaaaaagtc

atctatctacaaagcccttactttattccagaccaatcatatcataatgcattaaaaatg
gctgctaataagcggcgttgaaagtaaaccttatgataccgtgtaaacctgatcatccattc
gtttattgggtacatcttcaaatgcagctgatttattggatagcggaggttaatatattac
acttatcaaaatggatttattcattctaaaatattaatgattgatgatgaaatttcttca
5 attggtagtgcacacatggactttagaagctttgaactgaatttcgaagtgaatgcatt
atata

Sequence 1000

MNFGFLGTILTILLVVGFITNVVLA FVII FLERDRRTASSTWAWLFVLFVLPVIGFILYL
10 FLGRTVSKKKMEKNNNGDELHAFEDLVQDQIDSFDKHNYGYINDQVIKHRDLIRMLLMKQD
AFLTENNKIDLFDTGHLKLYEKVLEDIYNAQDYIHLEYTTFELDGLGKRILDALETKLKEG
LEVKLLYDDVGSKKVRLSKFKHFRALGGEVEAFFPSKVPLINFRMNNRNRHKIIIDGQI
GYVGGFNVGDDYLGLGKLG YWRDTHTRVQGECDALQLRFILDWNSQSHRPQFKFDQKYF
PKKNGDKGNAAIQIASSGPAFDLHQIEYGYTKMIMS AKKSIYLSQSPYFIPDQSYINALKM
15 AANSVGVEVNLMPCKPDHPFVYWATFSNAADLLDSGVNIYTYQNGFIH SKILMIDDEISS
IGSANMDFRSFELNFEVNAFIX

Sequence 1001

Contig_0542_pos_1002_1676,
20 putative peptide of unknown function
atgttgattatctttactgctttaatgattattgctaatttttactatatatctttttaa
aaaattggcttttactagtactcctattaggatgtgtacttgtatatgtagggtatgtg
tattttcataaagtaagaggactactatcttttggataggaaccttattaattgctttt
acacttttgtctaataagtacacgataattattctattttatatttttaatagtagtcac
25 atacgttatttggtttataagtttagacctttaaagtgttgcacagatgaagaaatc
acatcacccatttttattaagcaaaaatggtttgggtgaacaacatacacaccagtgtatgta
tataaatgggaagacgtacagattcaacacggtataggagacatacacattgatagaca
aaagcggcaaatattaaggaacaaataccatagttgtgcgtcatattttaggtaaagta
caagtagttgtacctcttaattataatataaaattacatgcgactctcttctacggcact
30 gcttatgtgaacgataaatcttataagattgagaataaccatgttcaaattgaagaaaaa
acgaaagatgataattatactgttaattgtttacgtttcatcattcattggagacgtagag
gtgatttacagatga

Sequence 1002

35 MLIIFTALMIANFYIIFFEKIGFLLVLLLGCVLVYVGYVYFHKVRGLLSFWIGTLLIAF
TLLSNKYTIILFIFLIVVIIRYLVYKFRPLKVIATDEEITSPIFIKQKWFEQHTPVYV
YKWEDVQIQHGIGDIHIDMTKAANIKETNTIVVRHILGKVQVVVPLNYNINLHATLFYGT
AYVNDKSYKIENNHVQIEEKTDDNYTVNVYVSSFIGDVEVIYR*

40 Sequence 1003

Contig_0542_pos_1703_2719,
is similar to (with p-value 1.0e-42)
>gp:gp|U81487|LLU81487_1 Lactococcus lactis subsp. cremoris
MG1363 histidine kinase (llkinD) gene, complete cds. NID: g2
45 182993.
atgcttatattagtagtatatagtagtcttattgcttttttatttattgataaagtgtttgta
aatattatctttttcaggggatgttttatacacaaatatttggaaatcctgtttttcta
tttttaaatatttaattgttcttttatgtattatagttggatctgttttagcttataaa
attaatcaacaaaatgattggattatttcacaaatagaaagatcaatagaaggacaaaca
50 gtaggtatcaatgatcaaaatatacgaattatatacagaaacgatagatatttatcataca
ctagttccattaaatcaagaattacatcgacttagaatgaagactcaaaatttaactaat
gaaaactacaatattaatgatgtaaaagtcaaaaagattatcgaagatgagcgacaacga
cttgccagggaattacatgattctgttagtcaacaattatttgcgtgcgagcatgatgcta
tcggcgataaaagatcgaaattagaaccacctttaaatacaacagataaccaattcttgaa
55 aaaaagggtcaagactcacaacttgaaatgagagctttgttattacatttaagaccgata
ggtttaaaagataagtctttagggtgaaggaattaaagatttagtcatcgatttacaaaag
aaagtaccaatgaaagtgtgcatgaaattcaagattttgaagtgccaaaaggcattgaa
gatcactgttcagaattacacaagaagctatttcaaatacattgagacattcaaatggt
acaaaagtaactgtggaattatttaatacaagaggattatcttttactaagaattcaagat

aatggaaaagggtttaatgtagatgaaaaatttgaacaaagttatggtttgaaaaatatg
 cgagaacgagcgcttagaaattgggtgcgacgtttcatattgtatctttacctgattcaggt
 acacgaattgaagttaaggcaccattgaataaggaggagaattcaagtggcgattaa

- 5 Sequence 1004
 MLILVYSMLIAFLFIDKVFVNIIFFQGMFYTQIFGIPVFLFLNLLIVLLCIIVGSVLAYK
 INQNDWIISQIERSIEGQTVGINDQNIELYTETIDIYHTLVPLNQELHRLRMKTQNLN
 ENYNINDVKVKKIIEDERQRLARELHDSVSQQLFAASMMLSAIKESKLEPPLNQIPILE
 10 KMQDSQLEMRALLHLRPIGLKDKSLGEGIKDLVIDLQKKVPMKVVEIQDFEVPKGIE
 DHLFRITQEASINTLRHSNGTKVTVELFNQEDYLLLRIQDNGKGFNVDEKFEQSYGLKNM
 RERALEIGATFHIVSLPDSGTRIEVKAPLNKEENSSGD*

Sequence 1005

Contig_0542_pos_2868_3338,

- 15 is similar to (with p-value 3.0e-24)
 >sp:sp|P55184|YXJL_BACSU HYPOTHETICAL TRANSCRIPTIONAL REGULA
 TOR IN GALE-PEPT INTERGENIC REGION. >gp:gp|X99339|BSGALE_5 B
 .subtilis orfs 1,2,3,4, pepT and gale genes. NID: g1429253.
 atggatttacttatggacgatatggatggtagaagcaactactgaaataaaaaaagat
 20 ttacctcaaattaaagtagtcatgttaacaagctttatagaggataaagaagttatcgt
 gcacttgattctggagtagatagttatattttaagacaacaagtgcagtgatatact
 gacgctgtgcgtaaaacgtatgaaggtgaatcagttttgaaccagaagtgttagtaaaa
 atgcgcaatcgtatgaaaaaacgtgccgagctttatgaaatggtgacagaaagagaaatg
 gagatcctattacttatagctaaaggatactctaaccaagagattgcaagcgctctcat
 25 atcaccatcaaaacagtaaaaactcatgtaagtaacatactaagtaaatagaagtacaa
 gatcgaaacacaagcagtaatatatgcgttcagcataatttaattcaataa

Sequence 1006

MDLLMDDMDGVEATTEIKKDLPIKVVMLTSFIEDKEVYRALDSGVDSYILKTTASDIA
 30 DAVRKTYEGESVFEPEVLVKMRNRMKKRAELYEMLTEREMEILLIAGKYSNQEIASH
 ITIKTVKTHVSNILSKLEVQDRTOAVIYAFQHNLIQ*

Sequence 1007

Contig_0542_pos_4575_3397,

- 35 is similar to (with p-value 1.0e-18)
 >gp:gp|AF071085|AF071085_2 Enterococcus faecalis strain OG1R
 F polysaccharide biosynthetic gene cluster, partial sequence
 . NID: g3608387.
 atgtcaaaaaaagagaaaaaacttctaataatcttaattcaatagaagataaagagcat
 40 aaaaagaataaaaaaataagaagttgaccgtacatatagaacctcaagaattccaatct
 aagaaacctaaaaaaagaatcaagttttttgtttcccggtgaataaaccagcaaaa
 tacaccgaaaactctaatttcttttcttacctgatttataggataggtaaagatgacgct
 gcaggtttagcagcacagatgacatatcattttgtattagcacttttcccaatgctaatt
 tttttacttacgctacttggtcaatttatcacgattgatgctaatacagattaatcaaaaa
 45 gtaagtcaatatgtccctgatcaagaaacagctagcatcggttggtggaattgttaaagat
 atctctgacactggcagtgaggatattttgtcagttgggttaatttttagctatttggtca
 gcatcaaatggaatgtccgctattattaactcatttaattgttgcttatgacggtgaggat
 tctcgaaacggtgtagtagttaaattattaagtattctatatacacttggttttaggtgca
 gtattttgttggtgctgtagtacttataacactaggtccagtcattaataaatttttattt
 50 ggaccactaggtattgataatcaaatggaatggatttttaatttagtacgaattgttatt
 ccattgattattattttcatcatatttactgtactttattcagttgcacctaatgttaaa
 acaaaattacgttctgttatttctgtgcgctattttcacttccattatttggttactaggg
 tcctttgcatcggttactatatttcaaaccttagtaactattcgaaaacatacgaagt
 ttagctggtattatcattttatttcttatgggtgtatatcacaaagctttattattattc
 55 ggtgctgaaatcaatgcaattattcaccaaagaaaagtcataagctggtcacacgcctgaa
 gaagccgctattaaacatgatgataacaatgaaaatcactataacgaaaatacgacttat
 gaatactatgaagatagcaagatgtagatatctctaataagatgacacgtataatatc
 aatcatcaatctaagaagaacatcacacaagcgactga

Sequence 1008

MSKKEKTSKYLSNIEDKEHKKNKKIEVDRTYIEPQEFQSKKPKKKNQVFFVSRLNKP
 YTENSNFYSYLIYRIGKDDAAGLAAQMTYHFVLALFPLIFLLTLLGQFITIDANQINQK
 VSQYVPDQETASIVGGIVKDISDTASGGILSVGLILAIWSASNGMSAIINSFNVAYDVED
 5 SRNGVVVKLLSILYTLVLGAVFVAVVLITLGPVINKFLFGLGIDNQIEWIFNLVRIVI
 PLIIIFIIFTVLYSVAPNVKTKLRVIPGAIFTSIIWLLGSFAFGYYISNFSNYSKTYGS
 LAGIIILFLWLYITSFIIIGAEINAIHQKVIAGHTPEEAAIKHDDNNENHYNENTTY
 EYEDSKDVDISNEDDTYNINHQSKEEHHTSD*

10 Sequence 1009

Contig_0543_pos_3868_3380,
 putative peptide of unknown function
 atgccaaaagtcacatcaagttaaggaaagatttgtgaaattaggggaccaacagtttaaa
 gcatthtgaattagatacgatacatattcattacgtgttgatgtgtgatgggtgtagat
 15 ttagcaatgaaacagcgcggtggaagatttgcagtgcgcaaacatggcatcaacaattt
 aaaacgattggcgctcatgctttttcaacaagataaacaattcatatatccactgatacat
 atacctaaatagatagcttaattctgggaaaatagctgtggttcaggagcggttctatc
 ggtgtgttagttaattatctaacagatcatgatattcaagattacctagttaaccaaccc
 ggaggcagattattgtctcatccagaaagtctggacaaaatgaatacacaacacgatt
 20 aagtgtcaagtttcaactgtcgcaacaggacaagcatatatagaacaggagacaatgacg
 caaatatga

Sequence 1010

MPKVHQVKERFVKLGDDQQFKAFEIRYDTHYVLMCDGVDLAMKQRVEDFVSAQTWHQQF
 25 KTIGVMLFQQDKQFIYPLIHIPKIDSLIWENSCGSGAASIGVLVNYLTDHDIQDYLVPNP
 GGSIIVSSRKSQNEYQTTIKQVSTVATGQAYIEQETMTQI*

Sequence 1011

Contig_0543_pos_2560_1283,
 putative peptide of unknown function
 atgggttggttagtgaccgggtcgctattcaacttgctcgactatgtcatttacatggagaa
 catatagttgatagtggtgagtcgcttcgcatcaacaaatctaagagagtcctttgat
 gcttatcaacgtgacggccttttttcagtaaatgactcaaaatgatgcacatcagtgcttt
 tcaggtaagtttacggttagacatttttttaaagatgttaaagatattactgaatattat
 35 gacgtggtgatttttagcatgtactgccgatgcgtatcgaccgatattacagcaattatct
 aagtcacacattaaagcgtattaaagcaaatcatcttggtctcaccaacattaggatcacat
 atgcttggttaagcaattactatcagatgttcaatgtgaaggatgaagtgttcatcttcc
 acttatctaggcgataaccgcaatatttgataaagcacacacacattgtgtcctaaccaca
 cgagttaaatcaaaattattcgtaggttcgactcaatctcagtcctatgacgttgtgtaag
 40 cttaagtcctttatttgactatttgaaatagaaatcaacacgatggacacaccactacat
 gcggagatacataatagttcactttatgtacacccaccattgtttatgaatcaattttca
 ttaaaggcggtattttgaaggacgaaagtaccagtatatgtatataagctatttccagag
 ggtccaatcacaatgaccttaatacacgaaatgcatgaatgtggcaagaaatgatgatg
 atattaaaaaattaaaggtaaccttcggtcaatcttctaaagtttatggtgaaagaaaac
 45 taccctatacgttatgagaccatgcgcgaagtagatattgaaaactttaaaaattacca
 gctattcatcaagagatctactttatgtgcgatatacagcaattttaatcgatccggtt
 tctaaccggacgatcaagggtcatattttgatttttctgccgtaccatacaaacatggt
 gatactgatgaacaaggagtcatacatataccacgcgatgccgagtgaagattattatcgt
 actttgataaattcaagcgattggaagagcattaaacgttgcaacaccgatgattgacaca
 50 ttgttattacgttatgaaaatactgttaacaatactgtgacacacatttacatcaacaa
 ctatcaaggcaattcgaattacatcattttaaacaggatttagcggttagtgacgaactac
 ttaactttttataaataa

Sequence 1012

55 MVGSGPVAIQLARLCHLHGEHIVDMVSRVHASTKSKRVFDAYQRDGFVSMTQNDHQC
 SGKFTVRHFFKDVKDITEYYDVVILACTADAYRPILQQLSKSTLKRIKQIILVSPTLGS
 MLVKQLLSDVQCEGEVISFSTYLGDTRIFDKAQPHCVLTTRVKSKLFVGSTQSQSMTLCK
 LKSLFDYLNIELTMDTPLHAEIHNSSLYVHPPLFMNQFSLKAVFEGTKVPVYVYKLFPE
 GPITMTLIHEMRMLMWQEMMILKKLVPSVNLKFMVKENYPIRYETMREVDIENFNPLP

AIHQEYLLVRYTAILIDPFSNPDDQGAYDFDSAVPYKHVD[~]TDEQGVIIHPRMPSEDYYR
TLIIQAIGRALNVATPMIDTLLRLRYENTVKQYCDTHLHQQLSRQFELHHFKQDLALVTNY
LTFYK*

```

5 Sequence 1013
  Contig_0543_pos_0_1232,
  is similar to (with p-value 0.0e+00)
>gp:gp|AF076683|AF076683_1 Staphylococcus aureus oligopeptid
e transporter putative substrate binding domain (opp-1A), ol
10 igo peptide transporter putative membrane permease domain (o
p-1B), oligopeptide transporter putative membrane permease d
omain (opp-1C), oligopeptide transporter putative ATPase dom
ain (opp-1D), and oligopeptide transporter putative ATPase d
omain (opp-1F) genes, complete cds; and unknown gene. NID: g
15 3800817.
  atgaataaactcacaaaactaagtacagtcatttttgtatctggaattatttttagccggt
  tgtggaataacaaagaactaacagagaaaaaagagaataaagtattatcatatacaact
  gtcaaagatattggagatatgaatccccatgtttatggagggttcaatgtcagcagaaagt
  atgatttatgagccggttagttcgcataccaaggatgggtattaagccattattagcaaaa
  20 aatgggacatttcacctgatggtaagacatatcggtttcatttaagggatgatgtatct
  tttcatgatggtacgaaatttgatgcagatgcagtgaaagaaaaacatcgatgcagtaaca
  caaataagaactacattctggttaagactttcaacactgatgtatggtcagaagt
  aaggataaagtatacagatacaactacattttgaaggaaacttatcaacctgctgttagcgaa
  ctgactatgccacgaccatacgtttttgtatcgctaaagatttttaaacacggcacaacc
  25 aaagatgggtgtgaaatcatttgacggtacaggaccatttaaaatgggtgaacacaaaaaa
  gatatatctgcagagtttaataaaaataatcaatattggggagaaaaggcaaaattaaat
  aaagtgaagcaaaagttaaacctgcaggagaaaacaacatttttatcaatgaaaaaagga
  gaaaccaactttgcttatacagatgatagaggtacagacagcttagataaagatagttta
  aaacaattaaaagaaaccggaagctaccaagtaaaacgtagccaggtatgaaatacaaaa
  30 atgcttgtgttataattctggttaagaaagatagtgacgtcagtgataaagcagtcagacaa
  gcattaggtcacatggtaaatagagataaaaatagctcaagatattttagacaaacaagaa
  aagccagccacacaactatttgctaaaaatgtgacagatataaactttaatttaccaaca
  agaacatatgataagaaaaaagcgcaagcggtatttagacaaggctggatgggtgctttca
  aaagatcgacaagttcgtcaaaaagagggcagaagatttgaatcttaagttgtattatgac
  35 aaagggtcttcagatcaaaaagacaagctgaattcttagaagcagaatttaagaagtta
  ggtgtcaactagatataaaacggagaaacgtT

```

Sequence 1014
40 MNKLTKLSTVIFVSGIILAGCGNNKELTEKKENKVLSTYTTVKDIGDMNPVHYGGSMSAES
MIYEPLVRNTKDGIKPLLAKKWDISPDKGTYTFHLRDDVSFHDGTFKFDADAVKKNIDAVQ
QNKKLHSLWRLSLTLIDDVVKVKDKYTIQLHLKEAYQPALAEAMPRPYFVFSPKDFKHGTT
KDGVSFSGDTGPFPMGEHKKDISAEFNKNQYWGEEKAMLNKVEAKVKPAGE'TTFLSMKKG
ETNFAYTDDRGTDSLKDLKQLKETGYSYQVKRSQAMNTKMLVNVSGKKASAVSDKAVRQ
ALGHMVNRDKIAQDILDKQEKPATQLFAKNVTDINFNLPTRTYDKKKAQALLDKAGWVLS
45 KDRQVRQKEGKDLNLKLYYDKGSSSQKEQAEFLEAEFKKLGVLQDINGETX

Sequence 1015
Contig_0545_pos_1330_1851,
putative peptide of unknown function

50 atgtcaaaaatcttaaacacacaattaactgggtatttttaatcggttgaaaaacaagag
ttggaattatcaaatggcagctcgaatgtctcattcaagcaattgggtggagaaggacatgct
tatatacaaggctacgatgattttaaattctatgagtcattcatattacaaagccatgaa
aaattagcgtctagcttaccacttgaagatttacaaaattttaacgatatagatacaaca
gatagggtactgttattttcaccatactacacttcggaagtgaaagtgatgtacttcaa
55 ctattgatttagatgtcgatttagtgcttatttgtaataaccctaaacgagatgatttt
cctaatacatttaattcattatgttaatttatcaacacctaggcccattgtttacacagaa
gattatgataaaatcattcaaccacattccgatggccttaaattatatttattatgatatt
tatactcaaatgattgaagatgactagaaacctaagatttataag

Sequence 1016

MSKILNTQLTGIFNRLEKQELDIQMAAQCLIQAIGGEGHVYIKGYDDLKFYESFILQSHE
KLASSLPLEDLQNFNDIDTTDRVLLFSPYYTSEVESDVLQLIDLVDLVLICNNPKRDDF
PNHLIHVNLSTPRPIVYTEDYDKIIQPHPMALNYIYYDIYTQMIEMTRDLDL*

5

Sequence 1017

Contig_0545_pos_6198_6629,

putative peptide of unknown function

atgatacaagggttaggctatttattgtccaatataacagattataaagaattaacgaat
10 ttagctcaaaatggagatcgtgatgccattgatttaaaagtaaaacatatttataaagat
actgaaccaccaattcctggagatttaacagcagcaaatgttgaaatgtattacatcac
ttagataatcagtttacatcagctaacaacttgccctctgcaattggcgctcgttggtgaa
gttataacaactatggctattacattagcacgtgaatataagactaagcacggtgtatat
atcggttcatcatttaataacaatcaattactacgtgaagttgttgaaaattacactgtt
15 ctaagaggatttaaacctactatattgagaatggtgctttttcaggcgctttaggagca
ctttacctctaa

Sequence 1018

MIQGLGYLLSNITDYKELTNLAQNGDRDAIDLKVKHIYKDTEPPPIPGDLTAANFGNVLHH
20 LDNQFTSANKLASAIGVGEVITTMAILTAREYKTKHVYIGSSFNNNQLLREVVENYTV
LRGFKPYYIENGAFSGALGALYL*

Sequence 1019

Contig_0545_pos_5531_4671,

25 putative peptide of unknown function

atgtttataaaaaaggattttgatgatattacagttcaagtatttgaagaaaaatataga
gatgcacttaaccaatttgaattaaagtgaacgacaacaaatatattcttcattgcctcaa
actgttttagatgatgcattaaaagatgaaaatcgaattgctaattgtagctttaaataaa
gaaggaaaagtagtgggttcttcgttttgcacgttattatcaacatgaaggttatgat
30 acaccaaacaatgttggtttatgtacgttcattgtcagttaatgaaaagtttcaaggccat
ggatatgggacaaaatgatgatgtttttaccagagtatgttcaagcattatttcctgat
tttacacattttatacttagtagtagacgctgaaaaccaaagtgttggaacgtttatgaa
cgtgcaggttttatgcatacagctacaaaagaagaaggacctattgggaaagaagactt
tattatttagatttagattcaaacatgtatcttctttaaggctaaaagagggggaagtc
35 acatataatgatgatattcacgtgattaatttgcctaaagatgatgtaaaggtaggcttt
attgcactagaacaaaatgataataaaatgaatatttctgcaatcgaagttaataagaaa
aataggaatgaggggaattgcagaaagtgcctttacgccaattaccaacgtatatacgtaaa
cagtttgaagacattgaagttttatcaattactttatatggcgaacgtaatgaattaaaa
ccattgtgcttgaatagtaattttgtagcaatagaggaaactgaggattatacacgcttt
40 gaaaaatatattaattattaa

Sequence 1020

MFIIKKDFDDITVQVFEEKYRDALNQFELSERQQIYSSLPQTVLDDALKDENRIANVALNK
EGKVVGFFVLHRYQHEGYDTPNNVVYVRSLSVNEKFQGHGYGKMMMFLPEYVQALFPD
45 FTHLYLVVDAENQSAWNVYERAGFMHTATKEEGPIGKERLYYLDLSKHVSSLRLKEGEV
TYNDDIHVINLLKDDVKVGFIALEQNDNKMNIASIEVNKKNRNEGIAESALRQLPTYIRK
QFEDIEVLSITLYGERNELKPLCLNSNFVAIEETEDYTRFEKYINY*

Sequence 1021

50 Contig_0545_pos_4560_4024,

is similar to (with p-value 1.0e-17)

>sp:sp|P12464|RPOE_BACSU DNA-DIRECTED RNA POLYMERASE DELTA S
UBUNIT (EC 2.7.7.6). >pir:pir|JT0302|JT0302 DNA-directed RNA
polymerase (EC 2.7.7.6) delta chain - Bacillus subtilis >gp
55 :gp|M21677|BACRPOE_1 B.subtilis RNA polymerase delta subunit
(rpoE) gene, complete cds. NID: g143455. >gp:gp|Z49782|BSDN
A320D_9 B.subtilis chromosomal DNA (region 320-321 degrees).
NID: g853752. >gp:gp|Z99123|BSUB0020_13 Bacillus subtilis c
omplete genome (section 20 of 21): from 3798401 to 4010550.

NID: g2636240.

atgaaaattcaagattacacaaaagaaatggttgatgagaaatcattcatcgatatggcc
tatactttattaaatgataaaacaaacacgatgaatttatatgatattattgatgaattt
aaatcttttaggcggatattgagtatgaagatattgaaaatcgaatcgtacaattctatacc
5 gatttaaacactgatggctggttttttaaatgtaggagaaaatctttggggtctacgtgat
tggtactctgtagatgatattgaggaaaaaatcgaccaacaattcaaaaattcgatatt
ctagatgacgaagatgaagaagatcaaaaccttaaaattattaggtgatgacgacgctgat
gaagatgacgatattcctgctcaaacagatgatcaagaaacattagacgagtcagataat
gatgaagatgatgttgaaatgaatgaagcagatatcgttattgatgaagacgaagacgaa
10 gatattgctgaaggtgaagaagaagcctttgaagacgccgaagactttaatgattaa

Sequence 1022

MKIQDYTKEMVDEKSFIDMAYTLNDKQTTMNLVDIIDEFKSLGGYEDIEDIENRIVQFYT
DLNTDGRFLNVGENLWGLRDWYSVDDIEEKIAPTIQKFDILDDDEEDQNLKLLGDDDDAD
15 EDDDIPAQTDDQETLDESNDNEDDVEMNEADIVIDEDEDEEDIAEGEEAFEDAEDFND*

Sequence 1023

Contig_0545_pos_3687_2080,
is similar to (with p-value 0.0e+00)
20 >sp:sp|P13242|PYRG_BACSU CTP SYNTHASE (EC 6.3.4.2) (UTP--AMM
ONIA LIGASE) (CTP SYNTHETASE). >pir:pir|A32354|SYBSTP CTP sy
nthase (EC 6.3.4.2) - Bacillus subtilis >gp:gp|M22039|BACSP0
OFA_2 Bacillus subtilis spo0F, CTP synthetase (ctrA), and f
ructose-bisphosphate aldolase (orfY-tsrf) genes, complete cds
25 . NID: g460910. >gp:gp|Z49782|BSDNA320D_10 B.subtilis chromo
somal DNA (region 320-321 degrees). NID: g853752. >gp:gp|Z99
123|BSUB0020_12 Bacillus subtilis complete genome (section 2
0 of 21): from 3798401 to 4010550. NID: g2636240.

atgacaaaagtttatttttgaacaggcggttgtgtcatcattaggaaggaataaca
30 gccgcttctctaggaagattacttaagatagaggacttaagttacaatacaaaaattc
gatccatatttaaatgtagaccaggcacaatgagtcggtatcaacatggtgaagtgttc
gttacagacgatggtgctgagactgatttagacttaggacattatgaacgttttatagat
attaatttaataaataattcaaatgttactgccgaaaagtatatcacatgtgttgaaa
aaagaacgcggtggtgattacttgggtggtactgtacaagttattcccatattacaaac
35 gaaattaaagaagattgctattagctggtgagagtactaatgcggatggttgaattact
gaaattggtggaacaacaggatgatagagtcctttacctttcttgaagccattcgtcaa
attagaagcgacttaggtcgtgaaaatgtaattgtatgtacattgtactttgctaccatat
attaaagctgctggggaaaatgaaaacaaacctacacagcacagtggttaaagaattacga
ggtctaggtattcaacctgatttaatagtagtacgtacagaatacgaatgacacaagat
40 ttgaaagacaaaatcgccctattttgtgatatacaaaaggaaagtgttatagaatgtaga
gatgcagattctctttatgaaattccgtttacaacttagtaagcaaatatggacgacatt
gttattcaacgtttacaattaaatgccaagtatgaaacgcaattggatgagtggaacat
ctattaaataccgttaataatttagatggtaaaattacaatcggttttagttggtaaatat
gtgagcttacaagatgcttatctatcagttgttgatcacttaagcatgctggttatcca
45 tttaaaaaagacgttggtggttaaaatggattgattcaagtgaggtcaatgatgataatggt
gaggcttatttatccgacgttgatggtatttttagttcctggtggatttggttgcagaca
agtgaaggtaaaattgcagctattcgttatgccggtgagaataacataccattctttggc
atttgtctaggaatgcaattggcaactggtgaatttgccgctcatgttttaggctatgaa
ggtgcgcattcagcagaattagatccaagtacaccatatccaattatagatttattacca
50 gaacaaaaagatattgaagatttaggtggaaccttaagacttggtctttatccttgccac
attaaagaaggtacattggcagagaaaatttataataaaaacgatattgaagaacgtcat
cgtcatagatatgaattcaataacgagtttagggaacaattagaaagtaacggtatggtta
ttttcaggtacaagtcagatggtcgttttagtagaaattattgaaatacctaaaaatgat
ttctttattgcatgtcaattccatcctgaattcttatcaagacctaatcgtccacagcct
55 atatttaaatcattttagaagcggcgttgaattaccaacaaaataa

Sequence 1024

MTKFI FVTGGVSSLGKGITAASLGRLLKDRGLKVTIQKFDPYLNVDPGTMSPYQHGEVF
VTDDGAETDLDLGHYERFIDINLNKYSNVTAGKVYSHVLKKERRGDYLGTVQVIPHITN

EIKERLLLAGESTNADVITEIGGTTGDIESLPFLEAIRQIRSDLGRENVMYVHCTLLPY
 IKAAGEMKTKPTQHSVKELRGLGIQPDILIVRTEYEMTQDLKDKIALFCDIKKESVIECR
 DADSLYEIPLQLSKQNMDDIVIQRLQLNAKYETQLDEWKHLLNTVNNLDGKITIGLVGKY
 VSLQDAYLSVVESLKHAGYPFKKD VVVKWIDSSEVNDDNVEAYLSDVDGILVPGGFGFRA
 5 SEGKIAAIRYARENNIPFFGICLGMQLATVEFARHVLGYEGAHS AELDPSTPYPIIDLLP
 EQKDIEDLGGTLRLGLYPCHIKEGT LAEKIYNKNDIEERHRHRYEFNNEFREQLSNGMV
 FSGTSPDGRLEIIEIPKNDFFIACQFHPEFLSRPNRPQPIFKSFVEAALNYQQK*

Sequence 1025

10 Contig_0545_pos_1097_252,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P13243|ALF1_BACSU PROBABLE FRUCTOSE-BISPHOSPHATE ALDO
 LASE 1 (EC 4.1.2.13). >pir:pir|S55426|D32354 fructose-bispho
 sphate aldolase (EC 4.1.2.13) - Bacillus subtilis >gp:gp|M22
 15 039|BACSP00FA_4 Bacillus subtilis spo0F, CTP synthetase (ct
 rA), and fructose-bisphosphate aldolase (orfY-tsr) genes, co
 mplete cds. NID: g460910. >gp:gp|Z49782|BSDNA320D_13 B.subti
 lis chromosomal DNA (region 320-321 degrees). NID: g853752.
 >gp:gp|Z99122|BSUB0019_209 Bacillus subtilis complete genome
 20 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 >gp:gp|Z99123|BSUB0020_9 Bacillus subtilis complete genome
 (section 20 of 21): from 3798401 to 4010550. NID: g2636240.
 atgaaagaaatgttaatcgatgcgaaagaaaacggttatgcggttggtcaatacaatctt
 aataacctcgaatttacacaagctattttagaagcgtctcaagaagagaatgcgccagtt
 25 atttttaggtgtttctgaaggggcagctcgttatatgagtggtttttatacagttgtgaaa
 atggtagaagggtttaatgcagtaacttaaacatcacaaatccagtagcaattcatttagac
 cagcgttcaagctttgaaaaatgtaaaagaagcaattgatgctggattcacatctgtaatg
 attgatgcacatctcatagtccttttgaaagaaatgttgaaatcacttctaaagtagttgag
 tatgctcatgatagaggcgtttctgtagaagctgaattaggtacagttggtggacaagaa
 30 gacgacgtagttgctgatggcggttatctatgcagaccctaaagaatgtcaagaattagta
 gaaaaaactggaattgatacttttagctccagcattaggttctgtacatggaccatataaa
 ggtgaacctaaattaggattttaagagatggaagaaattggtgcttcaactggattacct
 ttagtattacacgggtggtacaggtattccaactaaagatattcaaaaagctattccttat
 ggtactgctaaaaattaacgtgaatactgaaaatcaaattgcgtctgctaaagcagttcgt
 35 gaagtattaaacaacgacaaagatgtgtatgatccacgtaaatatttaggaccagcacgt
 gaagcaattaaagagacagttaaaggtaaaattagagaattcggtagcttctaatacgcgct
 aaataa

Sequence 1026

40 MKEMLIDAKENG YAVGQYNLNNLEFTQAILEASQEENAPVILGVSEGAARYMSGFYTVVK
 MVEGLMHDNLNITIPVAIHLHDHGSSFEKCKE AIDAGFTSVMIDASHSPFEENVEITSKVVE
 YAHDRGVSVEAELGT VGGQEDDVADGVIYADPKECQELVEKTGIDTLAPALGVS VHGPYK
 GEPKLGFKEMEEIGASTGLPLVLHGGTGIPTKDIQKAIPYGTAKINVNTENQIASAKAVR
 45 EVLNNDKDVYDPRKYLGPAREAIKETVKGKIREFGTSNRAK*

Sequence 1027

Contig_0546_pos_4340_4023,
 is similar to (with p-value 3.0e-30)
 >gp:gp|Z99119|BSUB0016_180 Bacillus subtilis complete genome
 50 (section 16 of 21): from 2997771 to 3213410. NID: g2635411.
 >gp:gp|U47861|BSU47861_2 Bacillus subtilis gbsAB operon, gl
 ycine betaine aldehyde dehydrogenase GbsA, alcohol dehydroge
 nase GbsB genes, complete cds. NID: g1524391.
 gtgtcaatttcacggttcccatTTTTtagtaaaaaagttgcggaagaaagtgaagaaatct
 55 ttctctgcaataaaatgttgctttcggtaccacgctgaattgttggttgacgatatcg
 tattctgttagttttttacacctgtactcatactaggtttactcatttgagttgtgtg
 cgcatttcatcaagtgtcatacttcttcaaaaaccataatgccatacaagttacctaca
 ctacggttgataccatacaaatccatggtttcaccgattgagttgataactaaatcttta
 gcttcttcgatataattga

Sequence 1028

VSISRSHFLVKKLRKKVKKSFSIAIKCCFRLPRVNCCLTISYSCSFFTPVLILGLLICSCC
RISSSVILPSKTIMPYKLPTLRLIPYKSMVSPIELITKSLASSIY*

5

Sequence 1029

Contig_0546_pos_3528_2038,

is similar to (with p-value 0.0e+00)

>sp:sp|P71016|DHAB_BACSU BETAINE ALDEHYDE DEHYDROGENASE (EC
1.2.1.8) (BADH). >gp:gp|Z99119|BSUB0016_179 Bacillus subtili
s complete genome (section 16 of 21): from 2997771 to 321341
0. NID: g2635411. >gp:gp|U47861|BSU47861_3 Bacillus subtilis
gbsAB operon, glycine betaine aldehyde dehydrogenase GbsA,
alcohol dehydrogenase GbsB genes, complete cds. NID: g152439

15

1.
atggaacttgtagataaattatcaaactcgatcaatatattgatggagaatgggttgaaagt
tcaaataaaaaacacaagagatattataaatccttacaatcaagaaacaatcttcactgta
gctgaaggaactaaagaagatggtgaaagagcaatttttagctgctagaagatctttcgaa
gacggtgaatggtcacttgaaacaagtgaagtcagaggtaaaaaagtgagagccgttgct
20 gataaaattaaagaaaaatagagaagagtttagctaaattagaaacattagacactggtaaa
actttagaagaatcctatgctgatattggatgatattcataatgtgtttatgtattttgct
ggttttagctgataaagatggcgggtgaaattatcaattcacctattcctaagtctgaaagt
aaagtagttaaagaacctgtaggtgtgttactcaaattacaccttggaactatccatta
cttcaagcatcttgaaaaattgcgccagcttttagcaacaggttgctcattagttatgaaa
25 ccaagtgaattactccgttaacaacaattcgtgtatttgaattgatggaggaagttggt
ttccctaagaagaacaatttaatttagtacttgggtgctggatcagaagtggcgacgtgatg
tcaggtcatgaagaagtcgatttagtttcatcaggtggtattgaaacaggaaaaacac
atcatgaaacaagcagcctaatacagtgactgacgttgcttagaattaggcgcaaaaat
cctaataattattttgatgacgtgattttgaattagctgtagaccaagcacttaattggt
30 ggatatttccacgctggtcaagtgtgctctgctggttcaagaatcttagttcacaatgat
attaaagataaattcgaaaaagctcttatcgatcgtgtaagcaaaaatcaaattaggtaac
ggttttgatcaagatactgaaatgggaccagttatctcaacagcacaccgagataaaaatt
gaagggttatatgaaagttgcgaaaaaagatggagcaacaattgcaattgggtgtaaacgc
cctgaacgtgaagacttacaagccggtattattctttgaacctactgtaattacagattgt
35 gatacatcaatgcgtattgttcaagaggaagtctttggaccagttgtgactgtagaagga
tttgctgacgaagaagaagctattcgcttagcaaatgattcaatttacggttttagcaggt
gctatatttactaaagatattggtaaagcacaacgtgttgcaataaattgaaacttggt
acggttttgattaacgatttccatccatactttgcacaagcgccatggggcggttacaaa
caatcaggtatcggtagagaattaggttaaagaaggattagaggaatatttagtaagtaaa
40 cacattcttacaataactaatccagaaccagtggttggttcagtaataaa

Sequence 1030

MELVDKLSNRQYIDGEWVSSNKNTRDIINPNQETIFTVAEGTKEDVERAILAARRSFE
DGEWSLETSEVRGKKVRAVDKIKENREELAKLETLDTGKLEESYADMDDIHNVMFYFA
45 GLADKDGGEIINSPIPAESKVVKEPVGVVQTITPWNYPQLQASWKIAPALATGCSLVMK
PSEITPLTTIRVFELMEEVGFPGKTINLVLAGSEVGDVMSGHEEVDLVSFTGGIETGKH
IMKQAAHVTDVALELGGKNPNII FDDADFE LAVDQALNGGYFHAGQVCSAGSRILVHND
IKDKFEKALIDRVSIKILGNGFDQDTEMGPVISTAHROKIEGYMEVAKKDGATIAIGGKR
PEREDLQAGLFFEPTVITDCDTSMRIVQEEVFGPVVTVEGFADEEEAIRLANDSIYGLAG
50 AIFTKDIGKAQVRANKLKLGTWINDFHPYFAQAPWGGYKQSGIGRELKKEGLEEYLVSK
HILTNTNPEPVDFWFSK*

Sequence 1031

Contig_0547_pos_495_1253,

is similar to (with p-value 4.0e-46)

>gp:gp|AF007865|AF007865_3 Bacillus licheniformis bacitracin
synthetase operon including bacitracin synthetase 1 (bacA),
2 (bacB) and 3 (bacC) genes, complete cds. NID: g2982193.
gtggcatattatgaagcatcgcaattaaaatcaacaggtcaattaaaagatatatttaagt

gaaacattacctgaatatatgatacctgtgcattttatgaaggtggatcgtatacctatc
 acgatgaatgggaaattagatgtgcgtgcattacctgaaattaatctaaagaataataga
 aattatgtagaaccacgtaacgatattgaacgcacagtttgccgtattttcgaagagatt
 ttacatggtgatcaggttaggtgttaaagataatttctttgaactaggtggacactctctt
 5 agagcaacatttagttgtaaacctgattgaagaaagggttaaaaaaacgtcttaaagtaggt
 gatttaatgaaatcgctactgtagagcaacttgacaacaaattgaagaactgcaaaat
 gatgtctatgaagtgattcccaaagcaaatgaatcgtatcaatatgatttaagtgcgtct
 caaaaaagtatgtatcttttatggaaggtcaatcctaagacacagtgataacattcca
 ttcttatggagattatcttctgaacttaattgttatgcaattgcaacgtgcattatctaag
 10 ttgattgaacgtcatgaaatattacgaacacaatatgtaattgatgacaatgaagttaaa
 caacgtattgcgacacatgtttcgctgattttgaagaggtaaccgacatctctaacgaa
 cgagcaagatattattcaatcatttatggaaccgtttga

Sequence 1032

15 VAYYEASQLKSTGQLKDILSETLPEYMI PVHFMKVDRIPI TMNGKLDVRALPEINLKNNR
 NYVEPRNDIERTVCRI FEEILHVDQVGKDNFFELGGHSLRATLVVNRIEERLKKRLKVG
 DLMKSPTVEQLGQQIEELQNDVYEVIPKANESYQYDLSASQKSMYLLWKVNPKDTVYNIP
 FLWRLSSELNVMQLQRALSKLIERHEILRTQYVIDDNEVKQRIATHVSPDFEEVTDISNE
 RARYYSIIYGTV*

20

Sequence 1033

Contig_0547_pos_1273_1977,
 is similar to (with p-value 2.0e-37)

>gp:gp|AF004835|AF004835_2 Brevibacillus brevis tyrocidine b
 iosynthesis operon, tyrocidine synthetase 1 (tycA), tyrocidine
 25 synthetase 2 (tycB), tyrocidine synthetase 3 (tycC), puta
 tive ABC-transporter TycD (tycD), putative ABC-transporter T
 yce (tycE) and putative thioesterase GrsT homolog (tycF) gen
 es, complete cds. NID: g2623770.

30 atgcgagttaaatatatacatggaccacaacaagattatatttatggatactcatcat
 agtattaatgatggatgatgtaaacacgattttactatctgatttgaacgctttataccaa
 gataaatcattacctgaacttaagcttcagtataaagattatagtgagtggatgggtgcac
 agagactttatctaacaacgtcacttttggttacagcaatttgaaaatcaggttccaata
 ttaaataatgcctacggattatcctagaccaagatttaaacaaccaacggtaatatgttg
 35 acgtttcattacaatcgtaaatcaaacagcaattgaaatcttatgtagaacaacatcaa
 gtgacagactttatgttctttgctagtgcattcatggtattattgcacaaatatacacgt
 caggacgatatcgctattggtagtgaatcagtgccgctactcatcgcgatactgaaaat
 atgttaggtatgtttgctaatacacttgatatcgtggctcgaccacatgatcaaaagaca
 tgggataattgatggctgagatgaaagagatgtgtctaggggcatatgaacatcaagaa
 40 tatccttttgaaagcttagtcattctatgaatggcctatttctccatccataacacgcgtc
 tactttaccagtttcttcattcgtacgatgatctttaatcattga

Sequence 1034

45 MRVKYIHGPPQDYLFMDTHHSINDGMSNTILLSDLNLYQDKSLPELKLQYKDYSEWMVH
 RDLKQRHFWLQQFENQVPILNMPTDYPRPSIKTTNGNMLTFHYNRQIKQQLKSYVEHQH
 VTDFFMFASAIMVLLHKYTRQDDIAIGSVISARTHRDTENMLGMFANTLVYRGRPHDQKT
 WDQLMAEMKEMCLGAYEHQYEPFESLVIYEWAYFSIHNSVYFTSFFIRTMIFNH*

Sequence 1035

50 Contig_0547_pos_5429_3171,
 is similar to (with p-value 0.0e+00)
 >sp:sp|006446|SECA STAAU PREPROTEIN TRANSLOCASE SECA SUBUNIT
 . >gp:gp|U97062|SAU97062_1 Staphylococcus aureus NCTC 8325 S
 eca (seca) gene, complete cds. NID: g2078389.

55 atgggtggtattgctatacataaaggtgatattgcagaaatgagaacaggtgaagggaaa
 acattgactgcaaccatgccgacgtattttgaatgcttttagctggtagaggtgtacatggt
 attacagtcaatgaatatctatcaagttcacaaagtgaagaaatggctgaactatataac
 tatcttggcttaactgtaggtttgaacttaaatagtaagtcaactgaagaaaaacgtgag
 gcttacgcacaagatatcatttatagtacgaataatgaacttggtttgattatcttaga

gataatatggtgaactatgctgaagagagagtaatgcgctcctctacattttgcaattatt
 gatgaggtcgattccatattgatcgacgaagcaagaacacctttaattatttctggtgaa
 gcggaaaaatctacttctttatatactcaagcaaatgtttttgcaaaaatgcttaaagcg
 gaagatgattataattatgatgaaaaaaccaaagctgtacatcttacagaacaaggtgca
 5 gataaagctgaacgtatggtcaaagtagataatctttatgatgttcaaaatgtggaagtg
 attagtcataattaatacagctttaagagctcatgttactttgcaacgcgatgttgattac
 atggtcgttgacggtgaagtattaattgttgaccaatttactggacgtacaatgcctgga
 cgtcgtttttctgaaggtttacaccaagcaattgaggctaaagaaggtgtagcaattcaa
 aatgagtcctaaaacgatggcatccattactttccaaaactatttcagaatgtataataag
 10 ttagcggggatgactggtacagcgaaaaccgaaggaagaatttcgtaatatctataat
 atgacagttacccaaattccaacaaacaaacctgttcaacgtaagataattcagactta
 atttatattagtcataaaaaggaaagtgttgatgcggtagtgtgaagatgttgtagaaaaacat
 aaaaaaggaacacccgtcttactaggtactgttgcgtgtgagacttctgaatatatttca
 aatttactaaaaaaacgtggtgtcagacatgacgtattaaacgctaaaaatcatgaacgc
 15 gaagctgaaatcgtttcaaacgcggggcaaaaagggtgcagttacaattgccacaaatatg
 gctggacgtggaacagatatataaacttgggtgatggtgttgaaaggttaggtggacttgc
 gttattggtactgagcgtcatgaatcaagacgtattgatgatcaattacgtggacgttca
 ggacgccaaaggtgatagaggagatagtcggttttacctatctttacaagatgaattaatg
 gtacgttttgggtcagaacgcttacagaaaaatgatgaaccgttttaggaatggatgattca
 20 acgccaatcgagtcgaaaaatggtatctcgagctgtagaatcagctcaaaaacgagtagaa
 ggtaataactttgacgcgcgtaaacgtatttctagaatacgtgaagttttacgtaagcaa
 cgtgaaattatttataatgagcgtaatgaaatcattgatagtgaagaaagttctcaagtc
 gttaacgcgatgttacgtttacattgcaacgtgcgattaatcattttattaatgaagaa
 gacgataatcctgactacacgccatttatcaattacgttaatgatgtgttcttgcaagaa
 25 ggagatcttcaagatacagaattaaaggtaaaagattcagaagatattttgaaattgta
 tggctctaaaattgaaaaagcatatgcacagcaacaagaacattaggagaccaaatgagt
 gaatttgacgggatgattttattacgttcaattgatacacattggactgatcatattgat
 acgatggatcaattgcgtcaaggtattcatttacgttcatatgcacaacaaaatccactt
 cgtgattatcaaaatgaaggtcatgaattatttgatatcatgatgcaaaatatcgaggaa
 30 gatacatgtaagtatatcttgaaatcagtggttcagtttgaaagatgatgtagaacgtgaa
 aatctaaaagctttggtgaagcaaaacatgtaactgctgaagatggcaagaaaaagca
 aagccccaaccgatgtgaaaaggtgatcaggtaggtagaatgatccatgcccatgtggt
 agtggttaaaaaatataaaaattgtcatgggaaagcgtaa

35 Sequence 1036
 MGGIAIHKGDIAEMRTGEGKTLTATMPTYLNALAGRGVHVITVNEYLSSSQSEEMAELYN
 YLGLTVGLNLNSKSTEEKREAYAQDITYSTNNELGFDYLRDNMNVYAEERVMRPLHFAII
 DEVDSLIDEARTPLIISGEAEKSTSLYQANVFAKMLKAEDDYNDEKTKAVHLTEQGA
 DKAERMFKVDNLYDVQNVEVISHINTALRAHVTLQRDQVDMVVDGEVLIVDQFTGRTPMG
 40 RRFSEGLHQAIIEAEGVAIQNESKTMASITFQNYFRMYNKLAMGTGAKTEEEEFNRIYN
 MVTVTQIPTNKPVRKQDNDLIYISQKGFDAVVEDVVEKHKKQPVLGTVAVETSEYIS
 NLLKKRGVRHVDVLNAKNHEREAEIVSNAGQKGAVTIATNMAGRGTDIKLGDCVEELGGLA
 VIGTERHESRRIDDQLRGRSGRQGDGRGDSRFYLSLQDELMVRFGERLQKMMNRLGMDDS
 TPIESKMVSRAVESAKRVEGNFDFARKRILEYDEVLRKQREIYNERNEIIDSEESSQV
 45 VNAMLRLSTLQRAINHFINEEDDNPDPYTPFINYNVNDVFLQEGDLQDTEIKGKDSIEDIFEIV
 WSKIEKAYAQOQETLGDQMSFERMILLRSIDTHWTDHIDTMDQLRQGIHLRSYAQQNPL
 RDYQNEGHELFDIMQNI EEDTCKYILKS VVQFEDDVEREKSFSFEAKHVTAEDGKEKA
 KPQPIVKGDQVGRNDPCPCGSGKKYKNCHGKA*

50 Sequence 1037
 Contig_0547_pos_2539_1892,
 is similar to (with p-value 1.0e-78)
 >gp:gp|299122|BSUB0019_26 Bacillus subtilis complete genome
 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 55 >gp:gp|AF013188|AF013188_1 Bacillus subtilis release factor
 2 (prfB) gene, complete cds. NID: g2331286. >gp:gp|AF017113|
 AF017113_1 Bacillus subtilis 300-304 degree genomic sequence
 . NID: g2618830.
 atgcttcttaggatgtatcaacgttactgtgaacaaaatggctttaagttgaaacgggtt

gattatcttccaggagatgaagcagggcggttaaaagtgtcacattacttattaaaggacat
aatgcttatgggtattttaaaaggcagaaaaaggtgttcacgttttagttagaatttcacct
ttcgattcatctggtagacgccatacttcttttgcacatgtgatgttattcctgatttt
aataatgatgaaattgaaatcgagattaaccagatgatatcacagtgagatacttttaga
5 gcttcagggcgctgggtggacaacatattaacaaaactgagctctgcaattagaattacacat
caccctacaggtattgtagtcaacaacccaaaatgaacgatctcaataaaaaatagagaa
gctgcaatgaaaatgttgaagtccaaactttatcaattaaagttagaagagcaagagcaa
gaaatggctgaaattcgagggcgaacaaaaagacattggatggggaagttagattcgttct
10 tacgtctttcatccatattcaatgattaaagatcatcgtacgaatgaagaaactggtaaa
gtagacgctgttatggatggagaaataggccattcatagatgactaa

Sequence 1038

MLLRMYQRYCEQNGFKVETVDYLPGEAGVKSVTLLIKGHNAYGYLKAKEGVHRLVRISP
FDSSGRRHTSFASCDVIPDFNNDIEIEINPDDITVDTFRAGAGGQHINKTESAIRITH
15 HPTGIVVNNQNERSQIKNREAAMKMLKSKLYQLKLEEQEQEMAEIRGEQKDIGWSQIRS
YVFHPYSMIKDHRTEETGKVDVMDGEIGPFIDD*

Sequence 1039

Contig_0548_pos_882_1841,
20 is similar to (with p-value 3.0e-70)
>sp:sp|P31114|GRC3_BACSU PROBABLE HEPTAPRENYL DIPHOSPHATE SY
NTHASE COMPONENT II (EC 2.5.1.30) (HEPPP SYNTHASE) (SPORE GE
RMINATION PROTEIN C3). >gp:gp|M80245|BACVARGNS_5 B.subtilis
dbpA, mtr(A,B), gerC(1-3), ndk, cheR, aro(B,E,F,H), trp(A-F)
25 , hisH, and tyrA genes, complete cds. NID: g143798. >gp:gp|Z
99115|BSUB0012_214 Bacillus subtilis complete genome (sectio
n 12 of 21): from 2195541 to 2409220. NID: g2634478.
gtggcaaagttaaacattaacaacgaaataaagaaagtagaaaagcgacttgaagaagca
attataagttctgatcaaacattacaagaagcctcattccatttactatcttcaggggga
30 aaaagagtttagaccgcttttgttattttaagtgggtcaatttggctctaacaacaaacct
tcagaagacacgtatcgtgtagcagtagctttagaactaattcacatggctaccttagtc
cacgatgatgtgatatagataaaaagtataaacgtagagggcgactcactatttcaaaaaaa
tgggaccaaagtacagctattttaacaggaaatttcttacttgctatggggtcaagcat
ttatctgaaatcagtgatactcgtgtccattcgaccatttctaaatcaattgttgatgtg
35 ttagaggagaactattccaatttcaagatcaatttaatagcaatcaacaattactaat
tacttacgtcgtatcaaccgtaaaacagcacttcttattcaactgtctacacaagttggt
gcgattacttccaatgcgtcaaatgacgttattcgtaaattaaaaatgatcggacattat
ataggtatgagtttccaaataatagatgatgtgctagattttactagttctgaaaagaaa
cttggtgaagcgggttggtagtgaccttatgaatgggtcatattacattacctgtactatta
40 gaaatgcgaaaaaataagacttttaagataaaaatttcacaacttaatcctgacagtcct
caacatgcctttgaaacttgataacaataattagacagtcggaaagcatagaacaatca
aaacaaataagtgaagattttaataaagcaatcaatttaatcgatgaattagaggat
ggtcctaataaagaactatttagaaagcttattaaaaaaatgggaagtcgaaataagtaa

45

Sequence 1040

VAKLNINNEIKVEKRLEEAISSDQTLQEASFHLLSSGGKRVPAFVILSQFGSNNKP
SEDYRVAVALELIHMATLVHDDVIDKSDKRRGRILTISKKWDQSTAILTGNFLLAMGLKH
LSEISDTRVHSTISKSIVDVCRGELFQFQDQFNSNQITINYLRINRKTALLIQLSTQVG
50 AITSNASNDVIRKLKMIGHYIGMSFQIIDVLDFTSSEKKLGKPVGS DLMNGHITLPVLL
EMRKNKTFKDKISQLNPDSPQHAFETCITIRQSEIEQSKQISEKYLKAINLIDELED
GPNKELFRKLIKMGSRNK*

Sequence 1041

Contig_0549_pos_673_1410,
55 is similar to (with p-value 5.0e-94)
>sp:sp|Q02142|LEU2_LACLA 3-ISOPROPYLMALATE DEHYDRATASE (EC 4
.2.1.33) (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI).
>pir:pir|S35134|S35134 probable 3-isopropylmalate deh

ydratase (EC 4.2.1.33) chain leuC - *Lactococcus lactis* subsp. *lactis* >gp:gp|U92974|LLU92974_16 *Lactococcus lactis* unknown gene, partial cds, and HisC (hisC), unknown, HisG (hisG), unknown, HisB (hisB), unknown, HisH (hish), HisA (hisA), His F (hisF), HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB), LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB), IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR) genes, complete cds. NID: g2565137.

atggaagcacgtatgacgatttggtaatatggctattgaagcaggagcaaagtatggttta
 atgcaacctgatgaaacaacctttaattacgtaaaaggctcgtccttatgctactgatttt
 gatagttctatggcgtggtggaaagaactttattctgatgatgcctattttgataaa
 gttattgaactttgatgttaacaaatttagaacctcaagtaacttggggaactaaccagaa
 atgggagttagtttttagtaaatccattcccagaaattaaaaatgcaaatgaccaacgtgct
 tatgactatatgggacttcacccagggtcaaaaagccgaagatataaaattagggttatgtt
 tttttagggttcacgtacgaatgcaagattatctgatcttattgaagcaagtcattatt
 aaaggacaacaagttcatccaaatattactgctattgtggttccgggttcaagaactggt
 aagaaggaagctgaagctctgggactagataaaattatttaaaagatgctggatttgagtg
 cgtgaaccaggatgttctatgtgcttaggtatgaatccagatcaagttcctgaaggagta
 catttgcattccacgagtaatcgcaattttgaaggaagacaaggcaaggcgctcgtaca
 catttggctatcccctgctatggctgctgctgctgcgattaatggtaaatcattgatgtt
 agaaagggtggtagtataa

Sequence 1042

MEARMTICNMAIEAGAKYGLMQPDETTFNIVKGRPYATDFDSSMAWWKELYSDDDAYFDK
 VIELDVTNLEPQVTWGTNPFEMGVSFSPFPEIKNANDQRAYDYMGLHPGQKAEDIKLGIV
 FLGSCNARLSDLIEASHIKGQQVHPNITAIIVPGSRTVKKEAEALGLDKLFDAGFEW
 REPGCSMCLGMNPDQVPEGVHCASTSNRNFEGRQKGARTHLVSPAMAAAAAINGKFIDV
 RKVVV*

Sequence 1043

Contig_0549_pos_1498_1980,
 is similar to (with p-value 6.0e-48)
 >sp:sp|Q02144|LEUD_LACLA 3-ISOPROPYLMALATE DEHYDRATASE SMALL
 SUBUNIT (EC 4.2.1.33) (ISOPROPYLMALATE ISOMERASE) (ALPHA-IP
 M ISOMERASE). >pir:pir|E36889|E36889 probable 3-isopropylmal
 ate dehydratase (EC 4.2.1.33) chain leuD - *Lactococcus lacti*
 s subsp. *lactis* >gp:gp|U92974|LLU92974_17 *Lactococcus lactis*
 unknown gene, partial cds, and HisC (hisC), unknown, HisG (hisG),
 unknown, HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF),
 HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB), LeuC (leuC),
 LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB), IlvN, IlvC (ilvC),
 IlvA (ilvA), AldB (aldB) and aldR (aldR) genes, complete cds. NID: g2565137.

gtgcatcttaagcgggtctctaaatcaggcttttgaccttttgcttttgatgaatggcgt
 tacttacctgatggtagtataatcctgattttaatacctaataaaccaaatatcatggt
 gcgtcaattctaattactggagataaactttggttgtggttctagccgtgagcatgcagcg
 tgggccttaaaagattatggttttaacattattattgcaggaagtttttagtgacatcttt
 tacatgaattgtactaaaaacgcaatgttacctatatgtttaaatcagaaagaaagagaa
 catttagctcaatttgatgaaataactgttgatttacctaatacaaacagtgtctacgggtg
 tctcagtccttttcattttgatatagatgaaacctggaaaaataaattaatccatggctta
 gacgatattgctattactttacaatttgaaaatttaatagaaaaatacgaataaactttt
 taa

Sequence 1044

VHLKRVSKSGFGPFADFDEWRYLPDGSNDPDPNPNKPKYHGASILITGDNFGCGSSREHAA
 WALKDYGFNIIAGSFSDIFYMNCNKNAMLPICLNQKEREHLAQFDEITVDLPNQTVSTV
 SQSFHFDIDETWKNKLIHGLDDIAITLQFENLIEKYEKTF*

Sequence 1045

Contig_0549_pos_1989_3263,
is similar to (with p-value 0.0e+00)
>gp:gp|U92974|LLU92974_23 Lactococcus lactis unknown gene, p
artial cds, and HisC (hisC), unknown, HisG (hisG), unknown,
5 HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF),
HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB),
LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB),
IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR)
genes, complete cds. NID: g2565137.
10 atgacagtgcagtaagaactaaagtttcgacaaaagatatagatgaagcatatttacgt
ctaaaaaatatagtaaaaagaactcccttacaattcgaccattacttatctcaaaaatat
aattgtaattgtttattttaaaaagagaagatttacagtgggtacgatccctttaaattaaga
ggagcttataaatgctatttcagttattatccaatgaagaaaaataaaggattacttgc
gcaagtgcgtgaaatcatgctcaagggtgttgccttatactgccaaaaaactcaatttaaaa
15 gctgttattttcatgccagtaactacaccacgacaaaaaatcaatcaagtcaaattcttc
ggggatagtaacgtagaaatagtattaattggcgatacatttgatcactgcttagcacia
gctttaaactatacgaagcaacataaaatgaattttattgacccatttaataatgtatat
actattgcaggacaaggcacttttagctaaggaaatattaaatcaagctgaaaaagaggat
aaacatttgaattgtatttgcgtctataggtgggtggcggtcttatttcaggagtgaac
20 acatattttaagcacattccccccatactaaaattattgggtgtgaaccaaccggtgcc
agtagtatgtatcaatcagtcgttatcaaccatagtatagttactttagaaaatattgat
aagttgttgatggagcttcagtagcaagagttgggtgatattacctttgatattgcgaaa
gataaagtggatgattatgttcaagttgacgaaggagctgtttgctccacaattctggat
atgtactcctaaacaagcgattgttgcgtgaaccagctgggtgctttaagtgtgaagtcctta
25 gaacaatataaaaaagcagattgaaaaataaaactattgtatgcataagtgagggaac
aatgatattaatcgaatgaaagaattgaggagcggtcccttctatttgaagaaatgaaa
cattactttatttttaatttccccacaaagacctgggtgctttaagagaatttgtcaatgat
gtcctcggaacctcaagacgatattacaaaatttgaatatttaagaaaacatcacaaaac
actggaactgttattataggtatacagctgaaacatcatgatgatctcattcagttaaaa
30 gatcgctgatgtcaatttgatccttctaattttatatcaatgaaaataaaatgttatat
tcattacttattttaa

Sequence 1046

MTVTVRVKVSTKDIDEAYLRLKNIVKETPLQFDHYLSQKYNVYKREDLQWVRSFKLR
35 GAYNAISVLSNEEKNGITCASAGNHAQGVAYTAKKLNKAVIFMPVTTTPRQKINQVKFF
GDSNVEIVLIGDTFDHCLAQALNYTKQHKMNFIDPFNNVYTIAGQGTLAKEILNQAEKED
KTFDYVFAAIGGGGLISGVSTYFKAHSPHTKIIGVEPTGASSMYQSVVINHSIVTLENID
KFVDGASVARVGDITFDIAKDKVDDYVQVDEGAVCSTILDMYSKQAIWAEPAGALSVSAL
EQYKKQIENKTIVCIVSGNNDINRMKEIERSLLFEEMKHIFILNFPQRPALREFVND
40 VLGPDQDITKFEYLKKTSTQNTGTVIIGIQLKHHDDLILQLKDRVCQFDPNSNIYINENKMLY
SLLI*

Sequence 1047

Contig_0550_pos_4007_5077,
45 putative peptide of unknown function
atgaatttaagatcactagatacaaaaagtagaggataataaacactttatctgatgataag
aaacaagcgcttaaaacaagaattgataagactaagcaaagtattgaccgacaaagaaat
attattatagatcaactcaatggtgctagtaataaaaaacaagcaaccgaagatatctta
aatagtgttttcagcaaaaaatgaagtagaagacataatgaaacgtattaaaacaaatggc
50 cgaagtaaatgaagatattgctaatacaaltgccaaagcaaattgatgggtcttgcattaact
tctagtgatgatattttaaaatcaatgttagatcaatctaaagataaagaaagtttaatt
aaacaattgttgacgacacgacttggtaatgatgaagcagatcgatttgctaaaaaattg
ttaagccaaaacttgctgaattctcaaatcgtagaacaattaaaacgtcatttcaatagt
caaggaacagactacagctgatgatatttgaatgggtgtgattaatgatgctaaagacaaa
55 agacaagcgattgaaacaatattacaaaaccgtatcaataaagacaaagctaaaattatc
gctgatgttattgcgcgtgtacaaaaggacaaatcagatatcatggatctcattcactct
gcgattgaaggcaaggcaaatgatttattagatatagaaaaacgagcaaaaacagctaag
aaagatttagaatatatttttagatcctataaagaatagaccatccttggttagatcgatt
aacaaggtgtcggtgatttcaattcaatatttgatagaccaagtttacttgataaactt

cactcaagaggatctattcttgataaattagatcattcggcaccggagaatggattatct
 ttagataataaaggtggccttttaagtgatctatttgacgacgatggtaatatctcatta
 ccagcgacaggtgaagtcatacaaacattggataaccagtggctgtgtactcatgtca
 ttaggtggagcgctcatctttatggcgcgtagaaaaaacacaaaaattaa

5

Sequence 1048

MNLRLSLDTKVEDNNTLSDDKKQALKQEIDKTKQSIDRQRNIIIDQLNGASNKKQATEDIL
 NSVFSKNEVEDIMKRIKTNGRSNEDIANQIAKQIDGLALTSSDDILKSMLDQSKDKESLI
 KQLLTTRLGNDEADRIAKKLLSQNLNSQIVEQLKRHFNSQGTATADDILNGVINDAKDK
 RQAIETILQTRINKDKAKIIADVIARVQKOKSDIMDLIHAIEGKANDLLDIEKRAKQAK
 KDLEYILDPIKNRPSLLDRINKGVGDSNSIFDRPSLLDKLHSRGSILDKLDHSAPENGSL
 LDNKGGLLSDLFDDDGNIISLPATGEVIKQHWIPVAVVLMISLGGALIFMARRKKHQ*

10

Sequence 1049

15 Contig_0550_pos_7700_10084,

is similar to (with p-value 5.0e-70)

>sp:sp|P49022|PIP_LACLA PHAGE INFECTION PROTEIN. >gp:gp|L146
 79|LACPIP_1 Lactococcus lactis pip and gerC2 genes, complete
 cds's, and rrg gene, 5' end of cds. NID: g308860.

20

atgaaaaacgcactaaaactttttatcacgatttaaaaaagagttgctaaaaacaccaggt
 gtatgggtcatcttagctgggttagcaattcttcttctcattctatgcatggtttaacctc
 tgggctatgtgggatccgtatgggtacacaggacatatcaaagttgccgtagtgatgaa
 gaccaaggtgaaaaagttcgtggtaagaatattaatgtaggaaataaaatggtcaaaact
 ttaaaaaagaatgatagttttgactggcaatttgtgagtagagaaaaagccgaccatgaa

25

attaagatgggaaaaatattatgcaggtatttatataccgaagaaattcacacatgaaatc
 actggtactttaagaaaaacatcctcaaaaggcggatatagattttaaagtaaatcagaag
 attaatgctgtagcagctaaagtaaccgatacgggacgtcgttctgtgattgataaagca
 aataaacaatttaacaaaaccgtagcaaccgctttactttctgaagctaataaagtcgga
 ctatcaattgaagataatgtacctacaatcaataaaattaagagtgctgtatatcaagct

30

aataattcattgcctaaaattaatcaatttgcagacaagattattgaactaaaataaacat
 caagacgatttggatgcttatgctaataattagaagtttaggaaagtataaagggaat
 gtattagacgctcaagaaaaacttaagtctgttaattcgtctattccggcgcttaatgaa
 agggctaaattgataacttgcacttgatagctacatgcctaatattgaaagaattttaaat
 gttgctgctaattgatgttccagcacaatttcttagaattaataggggtgtcgatattgca

35

agtgaaggtattgatgcagcgagtggtcagttaaatgatgcaaaaggttatttgactcaa
 gctaaagcgagagtgaggagactatcaagaagcagctggccgcgctcaagatgtgaacaac
 caagcaaatcaaaatctaagaaatcaaacatcaactacaccccaagcgctataaaatca
 tcgcatteggaaaggaagagtcattcaagcattaaaacagtcacctgtgagtcattcaggt

40

gagaatcaaccggtttatggtgataacattttatctaacagtgatgtaaaatcaatgaat
 acagctttaacgaagctttattatcattatctaatacaacagatcaacaagcacaagct
 acccaacaagacattaaagtcattaaaaaatatagcatatgggtgttatcgcttcagataaa
 ccatcagagttttaagaaccattaaaaaatataaatcacgcttagaaaacgcattctaag

45

tataatcaacaatttatagatatcttgtcagagttgaaaaaagtgaacatggtgatcta
 tctaataaagaaatgaagcaagtgaaagaagcaacaatagcattaatgataatttaaaaagt
 actaatcaattaatagatgcatttgtcaaatggtagctccggacaattagaagcagtcatt
 gtattacgtgacttacctaacttaataaaaaggttagatacattacgaaattacattaaa

50

aaagaacttaatacgttaatttactagctgtttctaataagattactgatcaacttaataaa
 ggtcaaaatacattatcgacaatccaatctaaattaaatactattaaccgagtgattaac
 gctggtcaagatatatttaaatagcggtaaaaagagaattgatacagattcaaactgcattg
 ccagcaatcgaaaacgcataataaatgcaatgcgaactgcacaagcttacttcccaaca

55

gctaaaaaagatgtcgcaagagctgcagactttgtacgtaattgacttgcctggattagag
 agagaattagctaattgaacacagctctgtaaaccaaaaaataccatctttatttagtcgt
 tatgataattgctgtagatttattaaacgagaacagcctcaagcaaaaagagcacttgct
 tcgcttgccgattttctcagaaaataaattgccagatgttgagaaagacttgaaaaaagca
 aataaaatcttcaaaaagttagataaagatgatgctgtagataagctaataagatacattg
 aaaaatgatttgaagaaacaggcagatatgttgctaacctatttaataaaaaaacgaca
 gatgtgttcccagtaaaaagactatggttctggatgacgccgttctatactgcattgtct
 atttgggttgaggattattaatggtcagcttattatccgatgtttcccttttccgacttg
 ataaccgtacaagccatctgtcttacttgttttgttttagataa

Sequence 1050

MKNALKLFITDLKRVAKTPGVWVILAGLAILPSFYAWENLWAMWDPYGHTGHIKVAVVNE
 DQGEKVRGKNINVGKMKVTLKKNDSTFDWQFVSREKADHEIKMGKYYAGIYIPKKFTHEI
 5 TGTLRKHHPQKADIDFKVNQKINAVAAKLTDGTSSSFVIDKANKQFNKTVATALLSEANKVG
 LSIEDNVPTINKIKSAVYQANNSLPKINQFADKIELNKHQDDLDAYANQFRSLGKYKGN
 VLDAQEKLNAVNSSIPALNERAKLILALDSYMPNIERILNVAANDVPAQFPRINRGVDIA
 SEGIDAAAGQLNDAGYLTQAKARVGDYQEAAGRAQDVNNQANQNLRNQTSTTPQSAIKS
 SHSEGKSHSSIKTVPVVSQSGENQPVYGDNLSNSDVKSMNTALTEALLSLSNQTDQQAQA
 10 TQQDIKSLKNIAYGVIASDKPSEFKEPLKNIKSRLENASKYNQOFIDILSELEKSEHVDL
 SNEIKQVKEANNSINDNLKSTNQLIDALSNGSSGQLEAVNVLRPLENLNKRDLTLRNYIK
 KELNRNLLAVSNEITDQLNKGQNTLSTIQSKLNTINRVINAGQDILNSGKKRIDTIQTAL
 PAIENAYINAMRTAQAYFPTAKKDVAKAADFVRNDLPGLERELANVTQSVNQKIPSLFSR
 YDNAVDLLNEKQPPQAKEALASLADFSENKLPDVEKDLKKANKIFKKLDKDDAVDKLIDTL
 15 KNDLKKQADIVANPINKKTTDVFPVKDYGSGMTPFYTALSIVWGGLLMVSLSDVSFSDL
 ITVQAICLTCFVFR*

Sequence 1051

Contig_0550_pos_7412_5445,
 20 is similar to (with p-value 6.0e-34)
 >sp:sp|P37710|ALYS_ENTFA AUTOLYSIN (EC 3.5.1.28) (N-ACETYLMU
 RAMOYL-L-ALANINE AMIDASE). >pir:pir|A38109|A38109 autolysin
 - Enterococcus faecalis >gp:gp|M58002|STRHYDROLA_1 Streptoco
 ccus faecalis bacterial cell wall hydrolase gene, complete c
 25 ds. NID: g153658.
 atgaagaaaaataaatttttagtataatttactatcgacggcgcttatcacgccaaccttc
 gctacacaaacagcttttgcgaagattcatctaataaaaaatacaaatcagataaaatg
 gaacaacatcaatcacaaaaagaaacatcaaaacaatctgaaaaagatgaatttaacaac
 gatgattctaaacacgattctgatgataaaaaagcacttctgacagcaaggacaaagac
 30 tctaataaaccattatcagctgactcaacacatcgtaactataaaatgaaagatgataat
 ttgattgatcaactttatgataattttaagtctcagtcagtagatttttctaaatactgg
 gaaccgaataaatacgaagacagtttttagtttaacgtcactcatacaaaatttatttgat
 ttgtattctgatataacagattacgaacagccacaaaagacaaagccattcttctaagac
 gaaaaagatcaagtagaccaagcagatcaggcaaaaacacatcacacatcaagaacaa
 35 tcacagtcgtctgctaaacaagatcaagaatcatcaaacgatgaaaaagaaaagacaact
 aacctcaagccgattctgacgtcagtgatttacttggagaaatggataaaagaagatcaa
 gaaggcgaaaacgtagatacaaaacaaaaatcaatcttctctgagcaacaacaaactcaa
 gcgaatgatgatagctcagaacgtaacaagaatatcttagtattacagattcagcatta
 gactctatatagatgaatatagtcaggacgctaagaaaacagaaaaagattacaataag
 40 agcaagaatacagaagtcacactaaaacatctcaaaagtataatgccgacaagaatccacaa
 ttaccaacagatgatgaattaaaacatcaatcaaaacctgcacaatcatttgaggatgac
 attaaacgctcaaatacacggttcaacaagctttttccaacaactacctgaattagacaat
 ggtgacttatcttctgattcatttaagtgtgttgacagtcagacacacgtagatttcatt
 caatcaattgctaaagatgcgcatcagatttgaaaaagaccaagatatatgcatcagtt
 45 atgattgctcaagctattttagaatctgactctgaaaaagttcacttgcacaatcacca
 aatcataacttggttggaatcaaagtgactacaaaggacaactctgtaacttttaatact
 ttagaagctgatagcagtaatacatatgttcagtatccaagcaggtttccgtaaaataccca
 agtactaaacaatctctgaagattatgcagatttaatacaacatggtatcgatggtaat
 ccgtcaattttataaaaccaacttggaagagtgaagctctatcatataaagatgctacttca
 50 catctgtcacgctcatagccacagatcctaattatttctaaaaaattaaatagtattatt
 aaacattatcatttaacatcttttgacaaaagaaaaatgcctaacatgaagaaatacaac
 aatcaataggtacggatgtctctggttaagtacttcaaaccatttactgaaacttccggt
 acatcaccttaccacatggccaatgtacttggatgtgtaccaccgtatgaatcaattt
 gatcatccatttctggtgactttagtgatgctcataattggaataatcgtgctgaaagt
 55 gaaggctatacggtaacgcacacacctaataatcatactcgagttgtgtttgaagctgga
 caattaggtgctgatacacagtatggtcatgttgcgtttcgttgaaaaagttaatgacgac
 ggttcaattgttatttctgaatcaaatgttaaaggattaggtgtcatttcattcagaact
 attgatgcagaagatgctcaagatttagattacattaaaggtaaatag

Sequence 1052

MKKNKFLVYLLSTALITPTFATQTAFÄEDSSNKNTNSDKMEQHQSQKETSQKSEKDEFNN
 DDSKHDSDDKKSTSDSKDKDSNKPLSADSTHRNYKMKDDNLVDQLYDNFKSQSVDFSKYW
 EPNKYEDSFSLTSLIQNLDFDSDITDYEQPQKTSHSSNDEKDQVDQADQAKQPSQHQQEQ
 5 SQSSAKQDQESSNDEKEKTTNHQADSVDSDLLGEMDKEDQEGENVDTNKNQSSSEQQQTQ
 ANDSSERNKKYSSITDSALDSILDEYSQDAKKTEKDYNKSKNTSHTKTSQSDNADKNPQ
 LPTDDELKHQSKPAQSFEDDIKRSNTRSTSLFQQLPELDNGDLSSDSFNVVDSQDTRDFI
 QSIKADAHQIGKQDDIYASVMIAQAILES DSGKSSLAQSPNHNLFGIKGDYKQSVTFNT
 LEADSSNHMFISIAGFRKYPSTKQSLDYADLIKHGIDGNPSIYKPTWKSEALSYKDATS
 10 HLSRSYATDPNYSKKNLSIKHYHLTSFDKEKMPNMKKYNKSIGTDVSGNDFKPFETETSG
 TSPYPHGQCTWYVYHRMNQFDASISGDLGDAHNWNNRAESEGYTVTHTPKNHTAVVFEAG
 QLGADTQYGHVAFVEKVNDGSIIVISESNVKGLGVISFRITDAEDAQDLDYIKGK*

Sequence 1053

15 Contig_0553_pos_3228_3920,
 putative peptide of unknown function
 atgcttaaaatagagagattaaccaaataatagacacgcaactgatatttaaagagata
 tcatgtacaattaacgaccagcacttactcataagtgggagagtggttggtgtaaattcc
 acattagqccaagattatcgctggcttagatagacgattatcagggcgaattatatcttaac
 20 gggcgcttacgtgaattcttatacgtctaaagagtggatgaagcacatccaatatgtacct
 caatatcaacgtgatacttttaaatcagcgtaaaacggtattagctacattattagaacca
 cttaagaattataaggtaaataaacagcggtatacatcaagcattgaagcagtgcttgat
 cagtgtaatttaccacacgatataacttaatacataaagtttcgacattaagtgggtggccaa
 tttcaacgcgctctggatagctaaagctttaattagaaccagagattctcatattggat
 25 gaagctacaaccaacttagatgtcattaatgaagaagctatacttcaaatgttgatttcc
 ttaagatgacacaattaatcattatttcacatgatacatagctcttaagccaatttgaa
 ggaattcatgactatcagtcgattccttttggttcttacgcaataagtcctcatctaatt
 ctggttaagggtacgagtgaaccatatcggttaa

Sequence 1054

30 MLKIERLTKYIDTQLIFKEISCTINDQHLLISGESGCGKSTLAKIIAGLDTDYQGELYLN
 GRLRESYTSKEWMKHIQYVPQYQRDTLNQRKTVLATLLEPLKNYKVNKQRYTSSIEAVLD
 QCNLPHDILNHKVSTLSGGQFQRVWIAKALILEPEILILDEATTNLDVINEEAILQMLIS
 LKMTQLIIISHDTYVLSQFEGIHQYSIPFGSYAISLHLILLRVRVNHIV*

Sequence 1055

35 Contig_0553_pos_5708_6910,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AB009635|AB009635_1 Staphylococcus aureus DNA for Fmt
 40 , complete cds. NID: g2696795.
 atgaaattaaataaattttaaatacgtattttttattatcgattatttaacttttagtcgtg
 tcaataggtatattaggtggtgaatggacaagacacctagaattaaaaaaacaaacgtta
 agtcaagaaagtggaaatacgaattatagaaaagagagataagactgttgagaaacct
 aaaaaataaagactaaatatgataaaaaagatcctacttccaaatcgataaacaaatat
 45 ttgaaaaaaactcaatttaattggaactgtagctgtatttgataatggaaaagttaaaatg
 aataaaggatatggttatcaagatatagagaaaggcaaaaagaacactgcaaatacaatg
 tatttaataggatcagcgcaaaaatttacaacaggtttaatgctgaagcaacttgaagtc
 gaaaaataaagtgaatttgcaagattcagtcactaaatatattccttggtttaaaacaaat
 aaagaaattacaattaaagatttaattgttacataaaaagtggaactatataaatatgaagct
 50 tcaactaatatcaaaaatttagaacaggctgttagagcaattcaagctcgaggtattgat
 gatacagtttatcataagcatcaatataatgatgctaattatttagtttttagctaaagtt
 attgaaaatgttactggaaaaccatatgttaaaaattattatgaacgattaggttaataaa
 tataatctcaaacatactgctttttatgacgagaaacctcttcaagtgagatggcaaaa
 ggctataagtttaaaaataataacttttcatctccttaaacctaataattatagatcaatat
 55 tatggagctggtaatttatatatgacgccacatgatatgggcaaatattttatcgtta
 caacaaaataaaatctttaatgcacgtcaaacctcgacctattttacatgaatttggaact
 caagaatatccagaagaatatagatatgggtttttacataactccgtatttaaatagagtc
 aacgggggtattcctttggtcaaattttactgtttactttaatgatcggtatattgtcatt
 ttaggagcgaatgtaagtaatacacctggattagtgagtaatgaagacaaaatgagacac

attttctataatattcttgaccagaaaaagccttataatacagcaggtgttaaagttgag
taa

Sequence 1056

5 MKLNKFKIVFFIIVLLTLVVSIGILGVEWTRHLELKKQTLSEQSGNTNYIEKRDKTVEKP
KKIKTKYDKKDPTSKSINKYLEKTQFNGTVAVFDNGKVKMNKGYGQDIEKGKNTANTM
YLIGSAQKFTTGLMLKQLEVENKVNLDQSVTKYIPWFKNKEITIKDLMLHKSGLYKYEA
STNIKNLEQAVRAIQARGIDDTVYHKHQYNDANYLVAKVIENVTKPKYVKNYERLGNK
10 YNLKHTAFYDEKPLQSEMAKGYKFKNNTFSFLKPNILDQYYGAGNLYMTPHDMGKLIYTL
QQNKIFNARQTRPILHEFGTQEYPEERYGYFITPYLNRVNGVFFGQIFTVYFNDRYIVI
LGTNVSNTPLVSNEDKMRHIFYNILDQKKPYNTAGVKVE*

Sequence 1057

Contig_0553_pos_11549_0,
15 putative peptide of unknown function
gtgttagaacgggaacacatttacgcgcatgcctttatttgaagaaaattatttatatg
tatgtgccccaaatcacatccactagctatgactgtacatccccgcgtatctcaatttaca
aatcaatcactatactgtctcgaaccaatgacaagctcaataaaaagtaaaattgattgaa
aagactaaggcacaagtagcaatgatttcagatatgaaactcgctcaacatattttgagt
20 cataataagggtattattttcttagtcaaaaattctttactatatgatcacgtaaaattgg
actaaaatcccttttaaatcatacagaattaaaacgaatgctatgtgtagttatgcgaaaa
gataacaagaaaaacgacattaatatagcatgga

Sequence 1058

25 VLERETHLRVMPLEENYMYVPKSHPLAMTVHPPLSQFTNQSLYCLEPMTSSIKSKLIE
KTKAQVRMISDMKLAQHILSHNKGFISSQNSLLYDHVNWTKIPLNHTELKRMLCVVMRK
DNKKNDINIAWX

Sequence 1059

30 Contig_0553_pos_10990_9044,
is similar to (with p-value 0.0e+00)
>sp:sp|P34956|QOX1_BACSU QUINOL OXIDASE POLYPEPTIDE I (EC 1.
9.3.-) (QUINOL OXIDASE AA3-600, SUBUNIT QOXB) (OXIDASE AA(3)
SUBUNIT 1). >pir:pir|B38129|B38129 quinol oxidase aa3-600 c
35 hain I - Bacillus subtilis >gp:gp|M86548|BACQOXA_2 Bacillus
subtilis AA3-600 quinol oxidase (QOXA, QOXB, QOXC, QOXD) gen
es, complete cds. NID: g143395. >gp:gp|X73124|BSGENR_39 B.su
btilis genomic region (325 to 333). NID: g413923. >gp:gp|299
123|BSUB0020_111 Bacillus subtilis complete genome (section
40 20 of 21): from 3798401 to 4010550. NID: g2636240.
atgattatctcagcacaattgctgcgccattcttagtcatcgcccttatagcagttata
tcttatttcaaattatggaaatatctatataaagaatgggtcacatccgtagaccataaa
aaaatcgggtatcatgtatttaatttctgcggtattaatgttcggttcgtgggtggtatcgat
gcgttaatgttacgtactcaattaacaattccagataacaaattcttggagcaaaccac
45 tataatgaagtatttactacgcacggcgtaattatgattatatttatggctatgccattt
atctttgggttatggaatggttatttccattacaacttgggtgcacgcgatggtgccttc
cctgtaataagaataacgttagtttctgctattcttcgctgggtatgattttattcaaccta
tcatttattgtagggtggatcaccagctgctgggtggactaactacgcaccacttgcgtggt
gaattcagtcagggtccaggtgttaactattttaaattgcaattcaaatatctgggtatc
50 ggatcggttaatgactgggtatcaacttcttgggtacgattctaagatgtaaaactccaaca
atgaagtttatgcaaatgccaatgttcagtgtaacaacattcattacaacattaatcggt
atattagcattcccagtggtcactgtagcacttgccttaatgactgctgtagaattttt
ggaactcagttcttactgtagcaaatggcggtatgccaatgctttgggcaaaccttcttc
tggttatgggggaccctgaagtttatcggtattttgcccagcattcggtatgtactca
55 gaaatcatccctacttttggccgtaaacgtttattcgggtcatcaaagtagatttgggca
actgcagggtatcgcatcttaagtttcttagtttgggttcaccatttcttactatgggt
aatgggtgcgttaattaactcatttcttctatctcaacaatgttaatcggtgttccaacc
ggagttaaactatttaactgggtgctcacattatacaaaggtagaattacatttgagtca
cctatgctattctcatttagcattcatccctaacttcttattaggagggttactgggtga

atgcttgcaatggcatcagctgactatcaatatcacaacacttatttcttagtagctcac
 ttccactatacattggttactgggtgtagtatttgcctgcttagctgggttaattcttctgg
 tatccaaaatgatgggctacaagttaaatgaaacattaaacaaatgggtgcttctggttc
 5 ttcgatgatcggtttaaactggttcttaccacaattcattctaggttttagatgggatg
 ccacgctcgtctatacacttacatgccttctgatgggttgggtggttactaaacttcattca
 actatcggtgcagttatgatggcaattggattcttattcctagttgcaagtatcggttat
 agtcatatcaaagctccacgtgaagctactggagataactgggatggacttggctgact
 ttagaatgggtctacagcatcagctattccacctaatacaactttgctatcactcctgat
 10 tggaatgactacgatacattcggttgatatgaaagaacatggctcgtcattatttagacaac
 cataactacaaagatattcatatgccaaacaatactccagtaggattctggatgggtata
 tttatgactattgggtggttcttcttaattcttcgaatctattgttccagcacttatctgt
 ttagcagggtatcttactactatgatttggagaagtttccaaattgatcatggttaccac
 atccctgcttcagaagttgcagaaactgaagctcggtttaagagaagctcgaattaaagaa
 agggaggctgtaagtcatgagtcatga

15

Sequence 1060

MIISAQIAAPFLVIGLIAVISYFKLWKYLYKEWFTSVDHKKIGIMYLISAVLMFVRGGID
 ALMLRTQLTIPDNKFLNHYNEVFTTHGVIMIIIFMAMPFIIFGLWNVVIPLQLGARDVAF
 PVMNNVSFWLFFAGMILFNLSFIVGGSPAAGWTNYAPLAGEFSPGPGVNYLIAIQISGI
 20 GSLMTGINFFVTILRCKTPTMKFMQMPMFVSVTFITTLIVILAFPVFTVALALMTADRI
 GTQFFTVANGGMPMLWANFFVWVGHPEVYIVILPAFGMYSEIIPFARKRLFGHQSMIWA
 TAGIAFLSFLVWVHHFFTMGNGALINSFISISTMLIGVPTGVKLFNWLTLTKGRITFES
 PMLFSLAFIPNELLGGVTGVMLAMASADYQYHNTYFLVAHFHYTLVTGVVFACLAGLIFW
 YPKMMGYKLNELTNKWCFFWMIGFNVCFLPQFILGLDGMPPRLTYTSPSGDWLLNFIS
 25 TIGAVLMAIGFLFLVASIVYSHIKAPREATGDNWDGLGRLEWSTASAIPKYNFAITPD
 WNDYDTFVDMKEHGRHYLDNHNYKDIHMPNNTPVGFWMGIFMTIGGFFLIFESIVPALIC
 LAGIFITMIWRSFQIDHGYHIPASEVAETEARLREARIKEREAVSHES*

Sequence 1061

30 Contig_0553_pos_8859_8449,
 is similar to (with p-value 4.0e-41)
 >sp:sp|P34958|QOX3_BACSU QUINOL OXIDASE POLYPEPTIDE III (EC
 1.9.3.-) (QUINOL OXIDASE AA3-600, SUBUNIT QOXC). >pir:pir|C3
 8129|C38129 quinol oxidase aa3-600 chain III - Bacillus subt
 35 ilis >gp:gp|M86548|BACQOXA_3 Bacillus subtilis AA3-600 quino
 l oxidase (QOXA, QOXB, QOXC, QOXD) genes, complete cds. NID:
 g143395. >gp:gp|X73124|BSGENR_40 B.subtilis genomic region
 (325 to 333). NID: g413923. >gp:gp|Z99123|BSUB0020_110 Bacil
 lus subtilis complete genome (section 20 of 21): from 379840
 40 1 to 4010550. NID: g2636240.

atgacttttgcattattaattagttcttatacttgtggtattgcaatttattacatgcga
 caagaaaaacaaaacttaatatgatgttttggatgattatcacagttatcctaggtcttgta
 ttcgtaggttttcgaaatttacgaattcgacactatgcttctgaaggtgtaaccaact
 attggctccttctggtctagtttcttatactactaggtacgcacggtgcacacgtatca
 45 ttaggtattgtttgggttatttgtttgtaattcaaactcggcactcgtggtttggattca
 tacaatgctcctaaattatttatagtaagttatactggcacttcttagatgttgtttgg
 gtcttcatcttactgccgtatatatgataggatgggtgtatagcggatga

Sequence 1062

50 MTFALLISSYTCGIAIYYMRQEKQNLMMFWMIIITVILGLVFGFEIYEFAYASEGVNPT
 IGSFWSFFILLGTHGAHVSLGIVVVICLLIQIGTRGLDSYNAPKLFIVSLYWHFLDVVV
 VFIFTAVYMIGMVYSG*

Sequence 1063

55 Contig_0553_pos_4902_3790,
 is similar to (with p-value 0.0e+00)
 >gp:gp|U071377|SEU71377_1 Staphylococcus epidermidis autolysi
 n AtlE and putative transcriptional regulator AtlR genes, co
 mplete cds. NID: g2267238.

atgaacaaatttataaaatattttttaatatattatcttttgggtctcctcgttggtcca
 attattttggctactcaattatatcaaagttcagaatcggcatcttgagtcattctcaaaac
 actaaagattctcaacgaaagtcactttaagagattcaaaagttgatcctgaaaaacaa
 cctatatcaattttattcttaggtatagacgataatgaaggtagagaaaaaacgggcaa
 5 agtgtagaacattctaggtcagatgctatgatattatctacttttaatacagaaaaagcat
 caaataagaatgcttagcatacctagagatactatcagttatatacctaaaagtggctat
 tacgataaaaataacacatgacgcatgcatatgggtggacctcttgctgctatggactcagtt
 gaagcaacaatgaatgtaccggtagattattatgtgcgtattaatatgaaagcctttggt
 gaagcagttgatgaattaggtgggtatattatgacgtaccatataacttaaatgaacct
 10 aacagtgatgatactggtagaattaaaataaaaaaaggataccaaaagctaaacggcgac
 caagcattagctgtagctcgaactagacaccatgattcagacctaaacgtgggtcaaaga
 caaatggaacttattaaaatattgttccaaaaagctcaaaatttaaaatctatagataaa
 cttgacaatggttattagattgttagggaaaaatgctaaacataatttaactcaaaaagaa
 attaaagtccttagccaaaatgtatcttgggtggtagtagtgaattaaaacatcacaaact
 15 aaaggtgaaggtgactacttaaatgatataactattaccaccaagcgtaaaaagtatt
 atggaatattcaaatcttttacgtaattgatttagatttatctaaaataacaaacaaaaac
 gatttcttagatcaaagagtcattaaacgatatgggtcactcgtacccttaacagaatta
 gatgaagacttattgcgtaagaacaaaaggaatcgactgatagtcattccttcaa
 attggcttaagacgtatgtatcatgtgaaataa

Sequence 1064
 MNKFIKYFLILLSFGLLVVPIIFATQLYQSSESFAESSQNTKDSQRKSTLRDSKVDPEKQ
 PISILFLGIDDNREGREKNGQSVEHSRSDAMILSTFNQKKHQIRMLSIPRDTISYIPKVG
 YDKITHAHAYGGPLAAMDSEATMNVVPDYVVRINMKAFVEAVDELGGIYYDVPYNLNEP
 25 NSDDTGRIKIKKGYQKLNGDQALAVARTRHDSDLKRGQRQMELIKILFQKAQNLKSIDK
 LDNVISIVGKNAKHNLTQKEIKSLAKMYLGGSTEIKTSQLKGKDDYLNIDYHYHPSVKSI
 MEYSNLLRNDLDSKITNKNDFLDQRIKRYGSLVPLTEDEDLRKNQKESTDSHEFLQ
 IGLRRMYHVK*

30 Sequence 1065
 Contig_0553_pos_0_2476,
 putative peptide of unknown function
 atgaatgcctatcaaattgaagaacttttttcacaagaaaatcttcaaaatgcagcangt
 tcaggccggtccaattcaatttcttggtaggttttgatggtgaagatagccatcataaccct
 35 gaaactcttttaccagtaaaatttatatgtaaaacctgagttaaaacatacaattgagtta
 tatcacgataatgaaaaacaagatagaaaggaattttcagtatcgaaacgagcgggcat
 ggtgttttccaagtaatgagtgaacgcttcataacactgtaggatcaggaatattacct
 tatcaacaagagatacgtatcaaaacttactagtaattgaaccaattaaagatagtgaatgg
 tctattacaggataccttaacacgcttacattacaaaacgctgtgggtagaacaaataat
 40 gctactgaaaaaaaacttagctcttgggtcatattgatccaggaaattatttcatcact
 gttaaagttgggtgataaaagtagaacaatttgaaattagatcaaaaaccaactccaccaaga
 atcattacaactgctaattgaattacgtggaaatcctaaccacaagcctgaaataagagta
 acagataataccaaatgatactactgctaaaaatcaaaacttgtgatggcggaaccgatgg
 gatcatgatccagaaataaatccatatactgtccctgaaaactacacagtagttgcagaa
 45 gcataccatgataatgatccaagtaaaaaatggggctttaacattccggtcatcagactac
 cttaaagatctaccattaaagcgggtgaattaaaggcaattgtttattacaatcaatagta
 caatcaaaacttttagtaatagcgttccggttagtagcgatacaacaccacctacaattaat
 gaaccagcaggactagttcataagattacaggggagatcatgtagaaattactctcca
 gtcactgataatactggcgggttcaggtttaagagatgtaaacgtcaatttacctcaagg
 50 tggacaaaaacctttacaatcaatcctaataataataactgagggtagcgttaagttaatt
 ggtaatatacctagtaattgaagcatataataacgacatatcatttcaatattactgcaacc
 gataattctggaaatacaacaaatccagctaaaacctttattttaaatgttggtgaagttg
 gctgatgatttaaatccagtcggattatctagagatcaactacaattagtgcagaccct
 tcttcattatctaattccgaacgagaagaggtaaaaagaaaaataagtgaagcaaatgct
 55 aataaagatactatttattacaaaataacccaataactcgctggagtaaacggcgatgtt
 acattttattatagagatgggtctgttagatgttattgatgctgaaaaatgtaatacacat
 gagccccgaaagaaaatccattttcagtgaaaatggtaatacaaaaataaaaaagaagcagta
 atcactattgctagaggacaaaactataccattgggtccaaacttaagaaaatatttctca
 ttaagtaattgggttcggatttacctaataagagatttcacctctatatcagctattggatct

ttaccttcatcgagtgaattagtcgactcaatgttggaattataactatagagttaat
 gctaaaaatgcttatcataagactcaacaagaacttaattttaaaacttaaaatagtagag
 gttaatgcacctactggtaataatcggtgatataagagtttagtacttataatttaactaat
 gatgaaatcaataaaatcaaacaagcattttaagcagctaattctggacttaatttaaac
 5 gataacgatatacactgttttcgaataactttgaccatagaaatgttagtagtgtagacagta
 actatacgttaaggcgatttgataaaagagtttcatcaaatctcaataatatgaatttc
 ttacgttgggttaataaaggatgattataccatttcgtggacttctagtaagattcaa
 ggtagaataacagatgggtgattagaatggtcaccagatcataaatcacttatttataaa
 tatgatgcaacattaggtagacaaataaataactaatgacgtgttaactttacttcaagca
 10 acagctaaaaactcaaatttacgttcaaataatcaatagtaatgaaaaacagtttagcagaa
 cgagggtctaattgggtattctaaatctataattagagatgatggcgagaaatcttattta
 cttaactcaaactcttcaagtatttagacttagtagaaccagataatgggttacgggtgga
 cgtcaagtcagtcattctaacgttatataatgaaaaaattcttctatcgtaaatggg
 caagttccagaagctaattggggcatccgcttttaattattgataaagttggttaagcta
 15 gcggcaataatggtta

Sequence 1066
 MNAYQIEELFSQENLQNAARSGRPIQFLVGFDVEDSHHNPETLLPVNLYVKPELKHTIEL
 YHDNEKQDRKEFSVSKRAGHGVFQVMSGTLHNTVGSGILPYQQEIRIKLTSNEPIKDSEW
 20 SITGYPNTLTQNAVGRTNNAATEKNLALVGHIDPGNYFITVKFGDKVEQFEIRSKPTPPR
 IITTANELRGPNPHKPEIRVTDIPNDTTAKIKLVMGGTDGDHDEINPYTVPENYTVVAE
 AYHDNDPSKNGVLTFRSSDYLDLPLSGELKAIVYNYQYVQSNFSNSVPFSSDTPPTIN
 EPAGLVHKYRGRDHVEITLPVTDNTGGSLRDVNVNLPQGWTKFTINPNNNTEGLKLI
 GNIPSNEAYNTTYHFNITATDNSGNTTNPACTFILNVGKLADDLNPVGLSRDQLQLVTD
 25 SSLSNSEREVEVKRISEANANIRSYLLQNNPILAGVNGDVTFFYRDGSDVIDAENVITY
 EPERKSI FSENGNTNKKAVITARGQNYTIGPNLRKYFSLNSGSDLPNRDFTSISAIGS
 LPSSSEISRLNVGNYNRYRVNAKNAYHKTQQELNLKLKIVEVNAPTGNRRVYRVSTYNLTN
 DEINKIKQAFKAANSGLNLNDNDITVSNNFDHRNVSSVTVTIRKGDLIKEFSSNLNMMNF
 LRWVNIRDDYTISWTSSKIQGRNTDGGLEWSPDHKS LIYKYDATLGRQINTNDVLTLLQA
 30 TAKNSNLRNINSNEKQLAERGSNGYSKSIIRDGEEKSYLLNSNPIQVLDLVEPDNGYGG
 RQVSHSNVIYNEKNSSIVNGQVPEANGASAFNIDKVVKANAAANNX

Sequence 1067
 Contig_0554_pos_1606_3477,
 35 is similar to (with p-value 0.0e+00)
 >sp:sp|P17922|SYFB_BACSU PHENYLALANYL-TRNA SYNTHETASE BETA C
 HAIN (EC 6.1.1.20) (PHENYLALANINE-- TRNA LIGASE BETA CHAIN)
 (PHERS). >pir:pir|S11731|YFBSB phenylalanine--trna ligase (E
 C 6.1.1.20) beta chain - Bacillus subtilis >gp:gp|X53057|BSP
 40 HEST_2 B. subtilis pheS and pheT genes for phenylalanyl-trna
 synthetase alpha and beta subunits. NID: g40052.
 atggttaggtactgcgtatgaagtcgcagctttatatcaaactaaaatgaataaacctcag
 ttaacaagcaatgaaagtcagaatctgctaaagatgaattaacaatagaagttaaaaat
 gaagataaagcaccttactatagtcacgtgttggtcatgacgtgactattggaccttct
 45 ccagtatggatgcagttccgatttaattaaagcgggaatacgtccaattaataatgtggta
 gatatttccaattatgtacttttagaatatggccaacctctacacatgtttgatcaagaa
 caaattggttcgcaatctatagaagttagacaagctaaaaaagatgagacaatgagaact
 ttagatggtgaagaacgtcgattgttagatactgacattgtcattacaaatggcaaagac
 cctattgcattaggaggtgttatgggaggagatttctctgaagtcactgaacaaacacga
 50 catgttgtagtagaaggggctatctttgatcctgtatctattcgacatacatcacgccgt
 ttaaatttaagaagtgaatcttcgagtcgatttgaaaaggggattgcaactgaatttgct
 gatgaagctgtagacagagcttggtattttacttgaaagatatgcttcaggaacagtttta
 aaagaccgagtttcgcatggagatttaggatcatttgtgacccaatagaaattactgct
 gacaagtaaacggtacaattgggttttaatttaactgatgaagaaatcattgatattttt
 55 gagcaattaggatttgacactgaaaaataaaaatgggtgaaattatcgtaagtgtccctca
 agacgtaaaagatatcttattaaagaagacttaatagaagaagtagcacgtatatatgga
 tacgatgacataccatcaacgctacctgtatttaagatgttacaagtgagagaactaaca
 gatcgacagttttaaagcgctactgttaagaaacacttgagggcgctgggctagaccaa
 gctattacttattcattggtatcaaaaaatcatgctaccgattttgcactacaaaatcgt

cctacaattgaactacttatgcctatgagtgaaacacataccacattacgtcaaagttaa
 ttaccgcatttaattgatgcagtatcatacaatggtgctcgtaaaaatacaaatgttaag
 ttatatgaaattggacgtgtcttcttttggaacgggtgaaggtgagttaccagatgaagta
 5 gaatacttgagtggtatattaactggagattttgtaataacacttggcaaggtaagaaa
 gagtcagttgatttctatttaactaagggtattgttgaacgtattgctgaaaagcttaatt
 cttcaattcgatttttagagctgggtcaaatagatgggttacatccaggaagaacagcaatt
 gtgtcacttaattggtaaagatattggcttcataagtgagctacaccctacgttagctgca
 aacaatgatttaaaagcgtacgtatgtatttgaacttaattatgatgcaatgatggaagtt
 10 tctgtgggatataattaattatgagcctatacctagatttccagggtgtaaacacgtgatatt
 gcattagaagttaatcatgaagttacttcatctgaattgttatccattattcatgagaat
 ggtgaagatattttaaatgatacactcgtatttggatgtatacaggggtgaacatttagaa
 aaagggaaaaaatctattgcaattagacttagttatctagatacagaaaacacacttacc
 gatgaacgtgtaaatgttgtgcatgataaaaattttagaagcacttaaaaagcatggtgca
 attattagataa

15

Sequence 1068

MVGTAYEVAALYQTKMKNPQLTSNESQESAKDELTIIEVKNEDEKAPYY SARVVHDTVIGPS
 PVWMQFRLIKAGIRPINNVVDISNYVLLLEYGQPLHMFDEQEQIGSQSIEVRQAKKDETMRT
 LDGEERRLLDITDIVITNGKDPIALGGVMGGDFSEVTEQTRHVVEGAIFDPVSI RHTSRR
 20 LNLRSESSSRFEKGIATFEVDEAVDRACYLLERYASGTVLKDRVSHGDLGSEFVTPIEITA
 DKVNRTIGFNLTDEEIIIDIFEQLGFDTENKNGEIIIVNVPSRRKDISIKEDLIEEVARIYG
 YDDIPSTLPVFKDVTSGELTDRQFKTRTVKETLEGAGLDQAITYSLVSKNHATDFALQNR
 PTIELLMPMSEAHNTLRQSLPHLIDAVSYNVARKNTNVKLYEIGRVFFGNGEGELPDEV
 EYLSGILTGFVNNTWQKKESVDFYLTKGIVERIAEKLNLQFDFRAGQIDGLHPGRTAI
 25 VSLNGKDIGFIFELHPTLAANNDLKRTYVFELNYDAMMEVSVGYINYEPIPRFPGVTRDI
 ALEVNHEVTSSSELLSIHENGEDILNDTLVFDVYEHEHLEKGGKKSIAIRLSYLDTENTLT
 DERVNVVHDKILEALKKHGAIR*

Sequence 1069

30 Contig_0554_pos_4982_5503,

putative peptide of unknown function

atgctcattgatatagttgttcttcttattatttgttactttatagtgatagggtttcgt
 agaggattttgggtatcgatattgcactttgcttcttcaattgtatctttatatattgcg
 tcacaacattatcaatctatttgcgcaacggttagttgtcttggccatttccgaaaacg
 35 gtggcggttgacatggtctatactatttcttattgatcatttgcataacagatttgaaaaa
 gtgatagcatttattataatatttgggtatgtgtaagcttatttggatctagttgttgtt
 acatttgataatataataacgtataaaaagatacatttagtaagtcggatcgcagtgct
 gttttgagtatcatagcgggtttttatatatttcaaaattggactttatttattatcgcta
 tatccgcattcatttatacagtaaccaattatctcaatcgctattaagtcgagttgtgatt
 40 gaacaaattccttatttatcacaatttattttaaattataa

Sequence 1070

MLIDIVLLIICYFIVIGFRRGIWLSILHFASSIVSLYIASQHYQSIAQRLVVFVFPFKT
 VAFDMVYTIPYDHLQYRFKVI AFIIIFGMCKLILYL VVVTFDNIITYKKIHLVSRISV
 45 VLSIIAVFIY LQIGLYLLSLYPHSFIQYQLSQSLLSRVVEIQIPYLSQFILNL*

Sequence 1071

Contig_0554_pos_4456_3530,

is similar to (with p-value 2.0e-34)

50 >sp:sp|O07874|RNH2_STRPN RIBONUCLEASE HII (EC 3.1.26.4) (RNA
 SE HII). >gp:gp|U93576|SPU93576_1 Streptococcus pneumoniae r
 ibonuclease HII (rnhB) gene, complete cds. NID: g2209338.
 atgggaaatgctgatacaaaactcacgtcaaaagaaattcaatcattgatgggtcaaaact
 acttttgagacgacgaagttacctaagggtatgaaagctcgtagagatatcaaaatact
 55 gttatcaatatctatagttcttggcaagtaagtgtttcaaggtaagaatgctgaacaactt
 gtagtgcaattgctaccaaataaacaatcaacaactggcaaacatacatcatcaaataca
 actagatttcaatataatcgttttcattgtatttgaagcgatgaagcaggcagtgggcgac
 tattttgggtccattgactgtatgtgcagcttatgtgagccaatcacatatcaaaatctta
 aaagaacttgggtgtagatgattcaaaaaactaagcgatactaaatcgtcgatcttgca

gaacagctcattacctttatcccgcattctttattaacattagataatgttaagtataac
 gaacgacaaagtctaggtggtctcaagttaaaatgaaagctgtcttacataatgaagct
 atcaaaaatgtgcttcaaaaaattgagcaagatcaactggattatattgttattgatcaa
 tttgcaaagcgagaagtttatcaacattatgcattatcagcattaccttttctgacaaa
 5 acaaaaatttgaaacaaaagggtgaatctaaatcactagcaatcgcggcagcaagcattatt
 tctcgttatgcatttgttaaacacatggaccacatctctaaaaaactccatatggaaata
 ccaaaaggagcaagtaacaaagtagatttaattgccgctaaagtcattcaaaaatgat
 attcaacaacttgatactatttcaaaaaaacattttaaaaacagagataaagcaattcat
 10 cttatgaatcaaaaatacaataaataa

Sequence 1072

MGNVVYKLTKEIQSLMAQTTFFETTKLPQGMKARTRYQNTVINIYSSGKVMFQGKNAEQL
 ASQLLPNKQSTTGKHTSSNTTSIQYNRFHCIGSDEAGSGDYFGPLTVCAAYVSQSHIKIL
 KELGVDDSKKLSDTKIVDLAEQLITFIPHSLLTLDNVKYNERQSLGWSQVKMKAVLHNEA
 15 IKNVLQKIEQDQLDYIVIDQFAKREVYQHYALSALPFPDKTKFETKGESKSLAIAAASII
 SRYAFVKHMDHISKKLHMEIPKGASNKVDLIAAKVIQKYDIQQLDTISKKHFKNRDKAIH
 LMNQKYNK*

Sequence 1073

20 Contig_0557_pos_329_844,
 is similar to (with p-value 2.0e-86)
 >pir:pir|D43258|D43258 galactose-6-phosphate isomerase subun
 it LacB - Streptococcus mutans
 atgaaaattgcaataggttgcgatcatattgttactgatacaaaaatggaagtttcacaa
 25 cacttaaaatcacaggacatgaagtgatagattgttgaaacttatgatttcacacgtaca
 cattatccgattttatgaaaaaaggtaggagaaaaaagttgcgagtggtgaagcagattta
 ggtgtatgtatttgggtactggtgttaggaattagtaatgctgcaaacaaagtaccagggt
 gttagaactgcttttagtttagagatatgacatcagcgctttattctaaagaagagttaaac
 gccaatgttgtaagttttggcggtaaagtagcaggtgaattatttattttcgacatcggt
 30 gatgcattcattgaggcagagtacaaacctactgaagaaaataaaaaatttaattgctaaa
 atcaatcatttagaagcacataacaatgaccaagctgatccacatttcttcgacgagttc
 ttgaaaaatggaataaaagggtgaatatcacgattaa

Sequence 1074

35 MKIAIGCDHIVTDTKMEVSQHLKSQGHEVIDVGTYDFTRTHYPIYGKKVGEKVASGEADL
 GVCICGTGVGISNAANKVPGVRTALVRDMTSALYSKEELNANVVSFGGKVAGELFIFDIV
 DAFIEAEYKPTENKKLIAKINHLEAHNNOQADPHFFDEFLEKWNKGEYHD*

Sequence 1075

40 Contig_0557_pos_1793_2770,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P11100|LACD_STAAU TAGATOSE 1,6-DIPHOSPHATE ALDOLASE (EC 4.1.-.-). >pir:pir|S04359|S04359 lacD protein - Staphylococcus aureus >gp:gp|X14827|SALACCD_2 Staphylococcus aureus lacC and lacD genes. NID: g46604.
 atgacaaaatcacacaaaaaagtgatcaattgagaaattaagtaatacaagaaggtatt
 atttcagcttttagcatttgatcaacgtggtgcattaaaaagaatgatggcagaacatcaa
 tctgaaacaccaacagttgaacaaatagaacaattaaaagtacttgtttctgaagaatta
 actcaatatgcgtcttcaattttatttagatccagaatatggtttaccagcatcagatgct
 50 cgaaataatgactgcgactattacttgcatagcaaaaaactggatatgatgtgaatgcg
 aaaggctggttgcagattgcttggtagaatggtctgcgaaacggttgaagagcaaggg
 gccaatgcagttaaatttttactttattatgatgtagatgacacagaagaataaacata
 caaaagaaagcatatattgaacgaattggttcagaatgtgttgccgaagatattcctttc
 ttcttggaagttttaacatatgacgacaataattcctgacaataaaaagtcagaattcgct
 55 aaagtttaagccacgtaaagttaatgaagcaatgaagttattctctgaagatcgtttaat
 gtggatgtacttaagttgaagtacctgtgaatatgaattttgtggaaggattttcagaa
 ggagaagttgtttataactaaagaagaagctgcacaacatttccgtgatcaagatgcagct
 actcacttaccatatatttatttaagtgcaaggtgtatcagcagaattgttccaagataca
 ttaaaatttgcgcatgattctggtgcgcaattcaatggtgttttatgtggacgtgccuca

tggtcaggagcagttaaggtatacattgaagaaggagagcaagctgccagagaatggttg
cgtacggtaggatttaagaatattgatgattgaatacagattgaaaacaacagctaca
tcatggaaaaacaataa

- 5 Sequence 1076
MTKSQQKVSSIEKLSNQEGIIISALAFDQRGALKRMMAEHQSETPTVEQIEQLKVLVSEEL
TQYASSILLDPEYGLPASDARNNDGLLLAYEKTGYDVNAKGRLPDCLVEWSAKRLKEQG
ANAVKFLLYDVDDTEEINIQQKAYIERIGSECV AEDIPFFLEVLTYYDDNIPDNKSAEFA
KVKPRKVNEAMKLFSEDRFNVDLKVEVPVNMNFVEGFSEGEVVYTKEEAAQHFRDQDAA
10 THLPYIYLSAGVSAELFQDTLKF AHDGSAQFNGVLCGRATWSGAVKVYIEEGEQ AAREWL
RTVGFKNIDDLNTVLKTTATSWKNK*

Sequence 1077

- Contig_0557_pos_2790_3104,
15 is similar to (with p-value 2.0e-29)
>sp:sp|P02909|PTLA_STAAU PTS SYSTEM, LACTOSE-SPECIFIC IIA CO
MPO NENT (EIIA-LAC) (LACTOSE- PERMEASE IIA COMPONENT) (PHOSPH
OTRANSFERASE ENZYME II, A COMPONENT) (EC 2.7.1.69) (EIII-LAC
) . >gp:gp|J03479|STALACS_1 S.aureus enzyme III-lac (lacF), e
20 nzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) g
enes, complete cds. NID: g153036.
atgaatagagatgaggtacaattactcggatttgaaattgttgccatgctggggatgca
cgttcaaaattattagaagcttttaaatgctgctaaagatagtgaaattgataaagcagaa
caacttgtagaggaagcgaatgaatgtattgctaataacacataaagcacaacacaaatcctt
25 ctgactcaagaggctaaaggcgaggatcgcataatagtcactatgattcatggtcaa
gaccatttaattgacaacattacttttaaaagatttaattgaagcatttaattgaattatac
aaaaaaggagctga

Sequence 1078

- 30 MNRDEVQLLGF EIVAYAGDARSKLLEALNAKDSEFDKAEQLVEEANE C IANAHKAQTNL
LAQ EAKGED IAYSITMIHQDHLMTTLLKDLMKHLIELYKKGS*

Sequence 1079

- Contig_0557_pos_3110_4858,
35 is similar to (with p-value 0.0e+00)
>sp:sp|P11162|PTLB_STAAU PTS SYSTEM, LACTOSE-SPECIFIC IIBC C
OMPO NENT (EIIBC-LAC) (LACTOSE- PERMEASE IIBC COMPONENT) (PHO
SPHOTRANSFERASE ENZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-
LAC). >pir:pir|B28474|B28474 phosphotransferase system enzym
40 e II (EC 2.7.1.69), lactose-specific, factor II - Staphyloco
ccus aureus >gp:gp|J03479|STALACS_2 S.aureus enzyme III-lac
(lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase
(lacG) genes, complete cds. NID: g153036.
atgaataaattaatagcatggatagaaaaaggaaagccattctttgaaaaaatatcacga
aatatttttaagagcgattcgtgatggatttattgctgctattccaattatcttattc
45 tcaagtatatatttttaattacctaattgtaccaaattgtgtttggttttacttgagtaaa
actatggaaggtatattgatgaaaccctataactatacaatgggaatagttggtttgctt
gtagcaggaaccacagctaaatcttttaactgattcttacaatcgaaaactagataaaacg
aatcagattaactttatttcgacaatgatggcagctatttgggatttttattcttagct
50 gctgatcctgttaagatgggtggattttcaagtgcatttatgggaacaaaagggtttattg
acagcctttatttctgcgtttattaccgtgattgtttataatttctttgtcaaaagaaat
attaccattaaaatgcctaaagaagtaccaccaaataatctcaagattttaagatatt
ttccctttatcagccgtaatttttaattttgtatgcttttagacttactttctagagcaata
gtccacacgaatgtagcaaatgcagattataaaagtatttgagccactatttactgcggca
55 gatggttggttggttggttaacactcatattcggtgcgtttgcgttcttctggtttgtaggt
attcatggaccttctattgttgaaaccagcgattgcagcaattacttatgcgaaccttgaa
acaaatttacacttaatacaagctggagaacatgctgataaagtaattacaccgggtaca
cagatgttcgtagcaactatgggaggaaccgggtgcaacattagttgttccatttatgttt
atgtggttaacaaaatcaaaaagaaataaaagcgataggtagagcatcagtcgtacctaca

5 ttcctttggtgtcaatgaaccatactttttggtgcaccactagtactaaatccggtattc
 tttataacctttttttttgcacctatagtaaatatggatttttaattttttggtgat
 gttttaaatatgaatagtttttagtatctttttaccttggactactcctgggccactcggg
 attgttatggggactggatttgcatttttggtcatttggtagcaatattactttattggt
 10 gttgatgtgattatttactatccattcttaaaagtatacgaatgaacaagtgcctgaagaa
 gaattaggaataaagaagcaataatgaattaaaagaaaaagtatcagcaaacctttgat
 acgaaaaaagccgatgctatttttagcaactgcaggggcaagtgaagcggatactgatgat
 acatcttcagttgatgaaacaacttctacatcctctacagatactattagtgaacaaaca
 aatgttttagttttatgtgcaggtggaggtacaagtgggtttactagctaattgctttaaat
 15 aaagctgctgaagagtatgaagtaccagtaaaagcagcagcaggtgggttatggtgcacat
 atggatattatgaagattatcaattaattatcttagcaccacaagttgcttcgaatttt
 gaagatattaaacaagatactgatcgcttaggaattaaattagccaaaactgaaggcgct
 caatatatcaagttaacaagagacgggtgaggcggttttagaatttgtaaaacaacaattt
 aacaattaa

15 Sequence 1080
 MNKLIWIEKGKPFEEKISRNIYLRAIRDGFIAAIPILFSSIFILITYVFPNVFGFTWSK
 TMEGILMKPYNYTMGIVGLLVAGTTAKSLTDSYNRKLDKTNQINFISTMMAAICGFLFLA
 20 ADPVKDGGFSSAFMGTKLLTAFISAFITVIVYNFFVKNRITIKMPKEVPPNISQVFKDI
 FPLSAVILILYALDLSRAIVHTNVANAVLKVFEPLFTAADGWIGVTLIFGAFAFFWVVG
 IHGPSIVEPAIAAITYANLETNLHLIQAGEHADKVITPGTQMEVATMGGTGATLVVPMFM
 MWLTKSKRNKAIGRASVVPFFGVNEPILFGAPLVLPVFFIFPIFAPIVNIWIKFFVD
 VLNMSFSIFLPWTPGGLGIVMGTFAFWSFVLAILLIVDVIIYYPFLLKVYDEQVLEE
 ELGNKEANNELKEKVSANFDTKKADAILATAGASEADTDDTSSVDETTSTSSDTISEQT
 25 NVLVLCAAGGTSGLLANALNKAEEYEVVPVKAAGGYGAHMDIMKDYQLIILAPQVASNF
 EDIKQDTRLGKILAKTEGAQYIKLTRDGEAALEFVKQGFNN*

Sequence 1081
 Contig_0557_pos_4874_6286,
 30 is similar to (with p-value 0.0e+00)
 >sp:sp|P11175|LACG_STAAU 6-PHOSPHO-BETA-GALACTOSIDASE (EC 3.
 2.1.85) (BETA-D-PHOSPHOGALACTOSIDE GALACTOHYDROLASE) (PGALAS
 E) (P-BETA-GAL) (PBG). >pir:pir|A27233|A27233 beta-galactosi
 35 dase (EC 3.2.1.23) - Staphylococcus aureus >gp:gp|J03479|STA
 LACS_3 S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE),
 and phospho-beta-galactosidase (lacG) genes, complete cds.
 NID: g153036.
 atgactaagaaattacctgatgacttttatttttggtaggcaaccgctgcttatcaagca
 gaaggagctactcagactgatggtaaaagggcggtgctgcttgggacacgtatttagaggag
 40 aattattggtacacagctgaaccagcaagtgatttttataacagatatcctgttgacttg
 gaattaagtgaacgctttggtgtaaatggtatacgtatctcaattgcttggtctcgtatt
 tttcctaaagggttacggtgaagtgaatcaaaaagggtgctgagttattatcataatcttttc
 aaagaatgtcataaacgtcatgttgaaaccttttgaacattacatcactttgacacacca
 gaggtacttcacaaagatggagacttcttaaatcgtaaaacaatagactattttgtagat
 45 tatgctgaattttgttttaagaatttccagaagtttaagtattggacaacattcaatgaa
 attgggcccgattggtgatggtaaatatttagttggttaaattccctccaggtatcaaatat
 gactttgaaaaagtattccaatctcatcataatatgatggttgacacgcacgtgctggt
 aaacttttttaagatgaaaattataaggaggagaaatagggtgtgcatgcattacctaca
 aaatatccatgatccatctaatacctgaagatgtgagagcagccgaacttgaagacatt
 50 attcataataaattttatttttagatgcaacataccttggttaagtactcacgtgaaacgatg
 gaaggagtacaacacatcttatctgtgaatggtggtcaattagagatttctgatgaagac
 tacaaaatttttagatgaagctaaggatttaaacgatttcttaggtattaattattatatg
 agtgactggatggtggttttgaaggcgaatctgaaataacacataatgccactgggtgat
 aaagggtggatctaagtatcaacttaaaagggtgtaggacaacgtgaatttgatggtgatt
 55 cctagaaccgattgggattggatgatttatccacaagggtttatgaccaaattatgcgt
 gtagtaaaagattatccgaattatcataagatttatattactgaaaatggattaggtat
 aaagatgtattcgacgaaaaaagaaaaaacagtacatgacgatgcacgaattgactatatt
 aaacagcatctaagtgtgatagcagatgcgattgcagatgggtgccaatgttaagggatac
 ttcttatggtctcttatggatgtattttcatgggtcaaatgggttatgaaaaaagatacggg

ttattctacgttgattttgaaacacaagaaagattccctaagaaaagtgcattattggtac
aaagaacttgcaagaaagtaagaaattaataa

Sequence 1082

5 MTKKLPPDDFIFGGATAAYQAEGATQTDGKGRVAWDTYLEENYWYTAEPASDFYNRYPVDL
ELSERFGVNGIRISIAWSRIFPKGYGEVNQKGVVEYHNLFKECHKRHVEPFVTLHHFOTP
EVLHKDGDFLNRKTIIDYFVDYAEFCFKEFPEVKYWTTFNEIGPIGDGQYLVGKFPPGIKY
DFEKVFQSHHNMVAHARAVKLFKIDENYKGEIGVVHALPTKYPYDPSNPEDVRAAELEDI
10 IHNKFILDATYLGKYSRETMEGVQHILSVNNGGLEISDEDYKILDEAKDLNDFLGINYIM
SDWMRGFESEITHNATGDKGGSKYQLKGVGQREFDQVPRDWDWMIYPQGLYDQIMR
VVKDYPNYHKIYITENGLYKDVDFEKEKTVHDDARIDYIKQHLSVIADAIADGANVKGY
FLWSLMDVFSWSNGYEKRYGLFYVDFETQERFPKKSAYWYKELAESKEIK*

Sequence 1083

15 Contig_0557_pos_6797_7423,
putative peptide of unknown function
gtgagtagtggaaggaaaagtatatattttgataaaaaattaagtgaagatgcagcaaacctt
attgtcaaaagtagaatttaaaagataataaaaaatggaaattttaagaaaatgcttattgg
attaagaagttctatcacactaaaaagtcattttggaattcaacaatttaattttgta
20 ggacattcaatgggaacatgtcattttgctttttacatgaaaaattatggggacgatcga
cattttgccacaacttaaaaaaggaagttaatatagcgggagtttataacgggattttgaat
atgaatgagaacgtgaatgaaattatcggttgataaacaggggaaaccaagtagaatgaat
gccgcataatcggaattgttatcactgcataagatttattgtggttaaggaaatagaagtt
ttaaatatctacggagatttagaagatggctcacattcagatggacgtgtgtcaaatagc
25 tcttctcaatcgcttcaatatattactaagaggtagcactaagtccttatcaagaaatgaaa
tttaaaggtgcaaaagcacacatagtcattacatgagaataaagatggtgcaaatgaa
atcatacaattcttatgggaaacttaa

Sequence 1084

30 VSSEGKVFYFDKKLSEDAANPIVKVEFKDNKNGNFKENAYWIKEVLSQLKSQFGIQQFNFV
GHSMGNMSFAFYMKNYGDDRHLPLQKKKEVNIAGVYNGILNMNENVNEIIVDKQGKPSRMN
AAYRQLLSLHKIYCGKEIEVLNIYGDLEDGSHSDGRVSNSSSQSLQYLLRGSTKSYQEMK
FKGAKAQHSQSLHENKDVANEIIQFLWET*

Sequence 1085

35 Contig_0557_pos_8545_7553,
is similar to (with p-value 2.0e-42)
>sp:sp|P39606|YWCH_BACSU HYPOTHETICAL 36.6 KD PROTEIN IN QOX
D-VPR INTERGENIC REGION. >pir:pir|S39699|S39699 hypothetical
40 protein - Bacillus subtilis >gp:gp|X73124|BSGENR_45 B.subti
lis genomic region (325 to 333). NID: g413923. >gp:gp|Z99123
|BSUB0020_105 Bacillus subtilis complete genome (section 20
of 21): from 3798401 to 4010550. NID: g2636240.
atgaagaggagatgtgtgttcattgagattaagtttacttgattatgtcccgttgttcgaa
45 gggcgtaccctcaaatgacgccctaaagcatagatttaaattagcccaacacgctgagaaa
cttgggtacttacgatactgggttcgagaacatcatcaagtttattctgtcgtttctagt
gcacctgaaataataatgatgtcgattttagaacacacacacacatcagagttggtagt
ggaggtgtgatgttaccacattatagtccttataaagtagctgagcaattttaaattatg
gaagcaagacacccccacgtatcgatatggctatcgacgttcgccaagctttaaaaat
50 gttaatgcagcactaaatgaaaacaaaaatgaaaaattaccattcaatactcagattact
gatttgcttaataacttcaataacgatacaactcaagaccatcggttttaaatcattatta
gctacacctatgggtacttcatttctcctaactatatatttttaggtatgagtaataagaagc
gcaaaaattagcttcagcgcgactacctttgttattgcagcaatgggacaactctgag
acagaccttcattgaagctataagcacttatagaaaatattttaaagcttatcatggtgaa
55 attaataatgcgaaccatattgtatttttagcaacttttgtggttaacagcttctaattta
tctagagttaaacaattgctacatacgttcaactttggttgatgcgtattaactattta
aatcaacctaaagattatccatcgattgaaacagcacagaacaagcattatagtcacaga
gaattagaaaagcttgaaaagatgaaatcgaaaatcatatacgaatgccaaatgatgttg
cggacaacttaccttacttcattcaacaatttaaagtggtgaaatcatcatcttacctc

atgtatttgggtgaagacgctagaatggaattaa

Sequence 1086

5 MKRRCVFMRLSLLDYVPLFEGRTPNDAKHSIKLAQHAKEKLGYLRYWVAEHHQVYSVVS
APEIIMMSILEHTQHIRVSGGVMLPHYSPYKVAEQFKIMEARHPQRIDMAIGRSPSFKN
VNAALNENKNEKLPFNTQITDLLKYFNNDTTQDHRFKSLLATPMVTSFPQLYLGMNSRS
AKLAAQRGLPFVIARMGQSETDLHEAISTYRKYFKAYHGEINNAKPYVILATFVVTASN
SRVKQLLHTLQLWLMRINYLNQPKSYPSIETAQNKHYSQRELEKLEKMKSKIIEYECQML
RNNLPYFINNLKWMKSSSYLMYLVKTLEWN*

10

Sequence 1087

Contig_0557_pos_2850_2530,
is similar to (with p-value 9.0e-35)
>sp:sp|P11100|LACD_STAAU TAGATOSE 1,6-DIPHOSPHATE ALDOLASE (EC 4.1.-.-). >pir:pir|S04359|S04359 lacD protein - Staphylococcus aureus >gp:gp|X14827|SALACCD_2 Staphylococcus aureus 1 acC and lacD genes. NID: g46604.
gtgcatccccagcataggcaacaatttcaaaccgagtaattgtacctcatctctattca
15 tttgaatatcctcccttacattatttgttttccatgatgtagctgttgttttcaatact
gtattcaaatcatcaatattcttaaatcctaccgtacgcaaccattctctggcagcttgc
20 tctccttcttcaatgtataccttaactgctcctgacctgtggcagctccacataaaaca
ccattgaattgcgaccagaatcatgcgcaattttaatgtatcttgaacaattctgct
gatacacctgcacttaataaa

25

Sequence 1088

VHPQHRQQFQIRVIVPHLYSFEYPPPLHYLFFHDVAVVFNTVFKSSIFLNPTVRNHSLAAC
SPSSMYTLTAPDHVARPHKTPLNCAPESCANFNVSWNNSADTPALK*

Sequence 1089

30 Contig_0558_pos_11130_10378,
is similar to (with p-value 0.0e+00)
>pir:pir|S19723|S19723 dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Staphylococcus aureus >gp:gp|X58434|SAPDHDNA_3 S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase. NID: g48871.
atgacacaacctgttaaaaaaggtatgaaagaaaaaggtatcgaaatcggtactgaagca
atggcaaaatctgcagaagaaactgaaaatgggtgtcaaagtaacttatgaggcaaaaggt
gaggaaacaaactatcgaaagctgattatgtattagttacagttggccgtcgccctaatact
40 gatgaattaggtatagaagaacttggtctgaaatttgctgatcggtgattactagaagtg
gacaaacaaagtcgtacttctattgaaaaatattctttgcgattggagatattgtacctgga
ttaccattagctcacaaagctagttatgaaggtaaagttgctgctgaagcgatagatggt
caagccgcagaggttagactatattgggtatgccagcagtttgctttacagaaccagaatta
gcacaagttgggttatactgaagctcaagcaaaagaagaaggtttatcaattaaagcttct
45 aaattcccttatgcagctaattggacgagctttatcattagatgatacaaatggttttgtt
aagttaattacacttaaaagaagatgatacgcttattggagcacaagttgttaggtactggc
gcatctgatattatctctgaattaggttttagctattgagtcaggtatgaatgctgaagat
atcgcatctaactgtacatgcacaccaacttttaggtgaaatgacaatggaagctgctgaa
aaagcaattgggttatccaattcatactatgtaa

50

Sequence 1090

MTQPVKKGMEKEGIEIVTEAMAKSAEETENGVKVITYEAKGEEQTIEADYVLTVGRRPNT
DELGLEELGLKFADRGLEVDKQSRSTSIENIFAIGDIVPLPLAHKASYEGKVAAEAIDG
QAAEVDYIGMPAVCFTEPELAQVGYTEAQAKEEGLSIKASKFPYAANGRALSLDDTNGFV
55 KLITLKEDDTLIGAQQVVGTSASDIISELGLAIESGMNAEDIALTVHAHPTLGEMTMEAAE
KAIGYPIHTM*

Sequence 1091

Contig_0558_pos_9659_9120,

putative peptide of unknown function

atggatattggatacaaaattacgtaatttaagaagaataaaaaatttgacacaagaggaa
 ttagcagagcgaaactgatttatcaaaaggatatatatcacaattgaaagtaatcatgct
 5 tcacctagtatggaaacatttttaaaatttaataagaagtacttggacttctgcaagtac
 ttttttaagaaccgtcagatgagaaggactttataagaagaaggaacagaccatttat
 gatgagtatgataaagggtatatcttgaactggctttagcgaattctaattgaatttgac
 atggaaccattaatcctaactttacgaccaaagcctcatataaaaaacttttaaccatct
 gaatcagatactttttatctattgttttaaatgggtgaagtatcacttcaattaggaaatcaa
 gtatataaagcttgtaaagaagatgtactttattttaagcgaaagataaacatcgctta
 10 tataacgaaacagataaagaagtgaagggttttaatcggtgccacagcttcatatttatag

Sequence 1092

MDIGYKLRNLRRIKNLTQEELAERTDLSKGYISQIESNHASPSMETFLNLIEVLGTSASD
 15 FFKEPSDEKVLKYKKKEQTIYDEYDKGYILNLVANSNEFDMEPLILTLRPNASYKNFKPS
 ESDTFIYCLNGEVSLQLGNQVYKACKEDVLYFKAKDKHRLYNETDKEVKVLIVATASYL*

Sequence 1093

20 Contig_0558_pos_9108_8014,
 is similar to (with p-value 5.0e-68)
 >gp:gp|AF077856|AF077856_1 Actinobacillus actinomycetemcomit
 ans putative polyamine transport operon, complete sequence.
 NID: g3341853.
 25 atgaatccattgctttcttttaagatgtcagtaagggtttgaagatgtacaaatacta
 aatgaaataatattgatattgaaccaggctattttatacactattaggtccctcagggt
 tgtggaaaaacaacaatttttaaaactcatagcaggatttgaatatcccgatagtgagat
 attatatataaagataaacctattggtaaaatgccaccgaataagcgtaaggtaaaatact
 gtattccaagactatgcattgtttccacatttaaatgtattcgacaatattgcatatgggt
 30 ttaaaattaaaaaaattaaagtaagtcagaaattaagcgtaagggttactgaagcacttcag
 ttggtgaaatgaagtggttatgaacataggcaaatacaagggtatgagtggtggacaaaaa
 caacgtgtagccatagcacgggcaattgttaatgagcctgaaatattattattagatgag
 tctttatccgcattagatttaaaattacgaactgaaatgcaatatttattgagagaactt
 caatcccggttttaggtataaaccttttatatttgaactcatgatcaagaagaggccttagca
 35 ttaagtgattatatttttgttatgaaagatggcaaaattcaacaatttggcacaccaata
 gatataacgatgaaccagttaaccgatttgttgctgattttataggagagtgccaacata
 gttcacggtacaatggttgaagattttgtcgtaaatattatgggtcaaaattttgattgt
 gtagatatgggaataaaaagaaaaataaaaaagttgaagttgtaattagaccggaagacatt
 tcacttgtttcacaaaatgatgggctatttaagccaaagttgattctatgctatttaga
 40 ggtgtacattatgaaattgttgtaagatagaaaagggtatgaatgggtaataacaatca
 acaaaaaaagctaattgtaggtagtgaagtaggtctgtattttgaaccagaagcaatacac
 atcatgggtaccaggtgaaactgaagaagaatttgataagcgaattgaaagttatgaggac
 tatcatcatgcataa

Sequence 1094

MNPLLSFKDVSKEFEDVQILNEINIDIEPGYFYTLGSPSGCKTTILKLIAGFEYPSDGD
 IIYKDKPIGKMPNKRKRVNTVFQDYALFPHLNVDNIAAYGLKLKLSKSEIKRKVTEALQ
 LVKLSGYEHRQIQGMSGGQKQORVAIARAIVNEPEILLDESLSALDLKLRTEMQYLLREL
 QSRIGITFIFVTHDQEEALALSDYIFVMKDGKIQFGTPIDIYDEPVNRFVADFIGESNI
 50 VHGTMVEDFVVNIYGQNFDCVDMGIKENKKVEVVIRPEDISLVSQNDGLFKAKVDSMLFR
 GVHYEICCKDRKGYEWVIQSTKKANVGSEVGLYFEPEAIHIMVPGETEEEFDKRIESYED
 YHHA*

Sequence 1095

55 Contig_0558_pos_7859_7218,
 is similar to (with p-value 1.0e-29)
 >sp:sp|P45170|POTB_HAEIN SPERMIDINE/PUTRESCINE TRANSPORT SYS
 TEM PERMEASE PROTEIN POTB. >pir:pir|A64118|A64118 spermidine
 /putrescine transport system permease protein (potB) homolog

- Haemophilus influenzae (strain Rd KW20) >gp:gp|U32813|U32813_11 Haemophilus influenzae Rd section 128 of 163 of the complete genome. NID: g1574796.

5 atgtttattgattcaatatggtatgccgctttaattactatgattaccttaataataagt
taccagctgcgtactttatttcttattcaagatttcaaaatatactgcttatgttgta
attatccctacttggattaatttacttcttaagacctatgcatttattggttgttgga
catgatggagttattaaccaagctctacatatatttcaaatacctaaattaaattgttg
tttacaagtggtgcatttttattgggtggcaggttatatttatatcccatattatgatttg
cctatatttaacagcatgaaagcaattcctaacaatatatttgcaagcctctaattgattg
10 ggcgcgagtacatttactacgtttcgtaaagtaatcggtcccttaacaagagaaggtatt
aaaacaggtgtgcaagtaacatttataccagctcttcactgtttttgattactaggttg
attgccgggaacaaagtaataatgtaggtacagcaattgaagaacagttcttaactata
caaaattatggattaggttccactatagcactttttctcattatttttatggccttttta
ctcattattacaaaatcaaaatcatctaattgggaaggggtga

15 Sequence 1096
MFIDSIWYAALITMITLIISYPAAFYFISYSRFQNILMLLIPTWINLLLKTYAFIGLLG
HDGVINQALHIFQIPKLNLLFTSGAFLLVASIYIYIPFMILPIFNSMKAIPNNILQASNDL
GASTFTTFRKVIVPLTREGIKTGVQVTFIPALSFLITRLIAGNKVINVGTAIEEQFLTII
20 QNYGLGSTIALFLIIFMAFLLIITKS KSSNGKG*

Sequence 1097
Contig_0558_pos_1609_272,
is similar to (with p-value 4.0e-63)
25 >sp:sp|P37536|YAAO_BACSU HYPOTHETICAL 53.2 KD PROTEIN IN XPA
C-ABRB INTERGENIC REGION. >gp:gp|D26185|BAC180K_91 B. subtilis
is DNA, 180 kilobase region of replication origin. NID: g467326.
>gp:gp|Z99104|BSUB0001_27 Bacillus subtilis complete genome (section 1 of 21): from 1 to 213080. NID: g2632267.

30 atgaaaaagaccaataattcaaaaattaaatcacttgatagagaaaaagctatctctatg
catgttcctggacataaaaaacatgacaatcggtacttaaataggcttgatttagcaatg
gatatgacagaaattactggattagatgatatgcattatcctgaaggaattattttagaa
agcatggagaatttttaggaacataaaaaactatgatgctttcttattagttaacggaacg
acttcagggtatattatcggttatccaagcgttttcgacaagaaaaggtaaatatttaatt
35 agtagaaatgttcataaatcagtattttcacggttttagacataacacaacaagcgaca
ataactaagacagatgtcagtaagaaaacgaatcaatatgtaaatccaaagataaatcaa
gataaaaatcaatattataaaacttgccatctgtacataccctaattattacggtgaaact
tttgatatttctcaatatatcaacaattacatcacagagggataccgatattagtagat
gaagcgcaggtgcacattttgggtttatggatttccagaatcctcaatgaattttaatt
40 gctgattacgttggcagtcataatcacaaaacactccctgcactaacaatgggatcagtt
atatatatacataaagatgcaccattaagacaacaagtcatagattatttaacttatttc
caaacgtcaagtccttcgtacctcattatgtctagtttagaattagcgaataaattttat
aaagaatatgattctacattatttgaccaacgaagaaagatgtaattgatttattagta
aatatgggatttacagttatagaaccagaggatccctttaaattgggtgtgagttttgag
45 ggtgttgagggttatgatgtgcaaaaatggtttgaggataaagaaatttatgtagaatta
gctgatgtatcaagtgttactcgttctccccctatggcatgaaggagataaatttcct
tttaagttggttgattgaaaaaattagagaaattaacgtgccaaaaaaatgtacgcgcgac
ataaaaacctcttaattttatgacgggttttagcgaatacaaaaactgttcattttcaaaat
acgaaagaagtgtctattaaaagggcagaaggttaaagtttttagcacaacatatcgttcca
50 tacctccaggtataccggtgatgttttaaggagaagtcgtgacgtctcatatgatagac
ttattaaataaatatgataaacaataatttaaagttgaaggttttaatacaaaaaaata
ttagttaaggtgaataa

Sequence 1098
55 MKRPIIQKLNHLIEKKAISMHPGHKNMTIGYLNRLDLAMDTEITGLDDMHYPEGIILE
SMENFRKHKNYDAFLVNGTTSGLSVIQAFTSRKGKYLISRNVHKS VFHGLDITQQQAT
ITKTDVSKKTNQYVNPKNQDKNQYKLAICTYPNYYGETFDISQYIKQLHHRGIPILVD
EAHGAHFGLYGFPESSMNFNADYVVQSYHKTLPALTMGSVIYIHKDAPLRQQVIDYLYF
QTSSPSYILMSSLELANKFYKEYDSTLFDQRRKMLIDLNVNMGFTVIEPEDPLKLVVFSF

GVEGYDVQKWFEDKEIYVELADMYQVLLVPLWHEGDKFPFKLLIEKIREINVPKKCTRD
 IKPLNFMTFSEYKTVHFQNTKEVSIKRAEGKVLAQHIVPYPPGIPVMFKGEVVTSHMID
 LLNKYDKQNIKVEGLNHKKILVKDE*

5 Sequence 1099
 Contig_0559_pos_867_1328,
 putative peptide of unknown function
 gtgaaaagtggcaaagcacgagcacatacaaatattgcggttgattaaagtattgggggaaa
 gctgatgaaacttacattattcctatgaataatagtttatcagttaccttagatagattt
 10 tatactgaaacaaaagtgcatttgaccctgattttactgaagattgccttatttttaaat
 ggtaatgaagtgaatgccaaagagaaaagattcaaaactatatgaatatagtgaga
 gatttggctggaaatcgtttgcagcgcaattgaaagtgaattatgtgccaaactgaa
 caatcaaaagaaaaacaagctaatagaacaagcaaaagcgcaaaatcttttgcgcgtgg
 agaaaagaagagcggttttgcatacgcacttactatcactttttctctattttcaaatgt
 15 tggatgcatagacgctcctttgactgtataagaagcaataa

Sequence 1100
 VKSGKARAHTNIALIKYWGADETYIIPMNNLSVTLDRFYTETKVTFDPDFTEDECLILN
 GNEVNAKEKEKIQNYMNIVRDLAGNRLHARIESENYVPTEQSKEKQANEQAKAQNLFARW
 20 RKEERFCYTTYHFFSIFKCMHRRSFDCIRSK*

Sequence 1101
 Contig_0559_pos_2914_3561,
 is similar to (with p-value 5.0e-52)
 25 >sp:P42423|YXDL_BACSU HYPOTHETICAL ABC TRANSPORTER ATP-BI
 NDING PROTEIN IN IDH 3'REGION. >gp:gp|D14399|BACIOLO_13 Baci
 llus subtilis 15 kb chromosome segment contains the iol oper
 on. NID: g709980. >gp:gp|Z99124|BSUB0021_68 Bacillus subtili
 s complete genome (section 21 of 21): from 3999281 to 421481
 30 4. NID: g2636442. >gp:gp|D45912|D45912_2 Bacillus subtilis g
 enome sequence between the iol and hut operon, partial and c
 omplete cds. NID: gl408482.
 atgggtccttctggatcaggtaaaacgactttactcaatgtgttaagttcaatagatact
 atttcagaaggaactgtggaagttgaaggcaagaaattaataaactgagccacaaagaa
 35 gtggcaaattttcgaaaacaacatctcgggttttatttttcaagattatagcggtttacc
 acattaacagtaaaagaaaatattatgctaccactctcagtaaaaaattccataaatat
 gaaatggaacaaaattataaagaagtggtgaggcattaggtatttataacctgggaaat
 aaatatccaagtgaatttctggcggtcagcaacaacgtacggcggcagcccgggcattc
 gtccataaaccacgattattttcgcagatgaacctactggcgcatagattctaaaagt
 40 gctcaagatttgttacaccgtctagaagatatgaataaacaatttaattcaaccattatg
 atggtgacacatgatccttcagccgctagttacgctgagagagtcattatgttgaaagac
 ggtgatatacactcagaatctaccagggtaacgattcaaaacaaacattttaccaagaa
 attatgaaacttcaaacgcattaggtggtgtcagtcagtcattttaa

45 Sequence 1102
 MGPSGSGKTTLLNVLSSIDTISEGTVEVEGKEINKLSHKEVANFRKQHLGFIQDYSVLP
 TLTVKENIMLPLSVQKFHKEYEMEONYKEVAEALGIYNLGNKYPSEISGGQQQRTAAARAF
 VHKPTIIFADEPTGALDSKSAQDLLHRLDMNKFNSTIMMVTHDPSAASYAERVIMLKD
 GDIHSEIYQGNDSKQTFYQEIMKLQ TALGGVSHDI*

50 Sequence 1103
 Contig_0559_pos_3857_0,
 putative peptide of unknown function
 atgctcaatatagaacaactagttatttttattgtaacaggaatttttaggcactttaatt
 55 ggtatttttgggttcaaaacttttacttgttatcgcttctaaattaatgaagttaaacaca
 catatctctattggcttgaaccccaagctatacttattactatcgtaattgttagctgtc
 gcttttttattgataatgatacaaaattacattttcttaaaaaaacacagcatttttagct
 ttgatgaaagacaattataccccggaagctacccaaaaacggataactacgtttgaagca
 atcgcgccgcattttaggaattataatgatagttttggatatttatgtctactgaaatg

tttgggtgtttttaagccttaacaactgctttgattacaccttttagcatacttttctta
actattgttgggtgctttcttattcttttagaagttctgtatcacttatttttaaacacta
aaacatattaaacatggctcggtgtaaatatcacagatgttgtctttacatcatctatcatg
cacagaatgaagaaaaatgcgatgtctctcacagttattgctatcatttcagctttcacg
5 gttagtattctttgcttcgcgcaattacacaatctaataactaatacaacttttagaaatg
acctctccagatgattttaataaagccagaataaaatagctgcgcaatttaaacataaa
ctagatcaaggaaatttaaaatatcatcagcggacttatgaagtaataatccaaaaaca
ttaagcgaccacgtcatgaagagtaaaaatggttctgatatgtctactaataacaatca
ctaagtatgaactcacatctcaaaggatcatgaagctaaaataacgaatatacaatcatca
10 acaggattaatagatattcattttaaatcataagattacagttaaaggaaaatctaataca
tctattatcggttaaagaca

Sequence 1104

MLNIEQLVFFIVTGILGLIGIFGSKLLVIAASKLMKLNTHISIGFEPQAILITIVMLAV
15 AFLIMIQNYIFLKHSILALMKDNYTPEATQKRITTFEAIGGILGIIMIVFGYYMSTEM
FGVFKALTALITPFSILFLTIVGAFLFFRSSVSLIFKTLKHKHGRVNI TDVVFSSIM
HRMKKNAMSLTVIAIISAFTVSILCFAAITQSNTNTLEMTSPDDFNISQNKIAAQFKHK
LDQGNLKYHQRTYEVINPKTSLDHVMKSKNGSDMSTNTTSLMMNSHLKGHEAKITNIQSS
20 TGLIDIHLNHNKITVKGSKQSIIVKDX

Sequence 1105

Contig_0559_pos_2535_2134,
putative peptide of unknown function

atggagaaaaggagtattaatatgaaaaagatattatgatcataagtatacttaccata
25 actgttactttaagtgcattgtggaggttctggaaaacaaaaagagccatctaaggaaagt
caaaaatctgataaatatgattatgtttattatgaaatatataatgatggagattctgaa
acgccaaatgttgagattaaatataaagataaaaaaggtaaatcacatatagaaaaagct
gatttagatcacgtgtatgaacatatactaggtgatggtaataaaaaaccatatattgta
aaggatgggaagaaaattcatgtatatcgaccaccatatatgatttatggtgatgatgat
30 gttgaaggcaagccgtttcgaaagatgaagttacgaagtaa

Sequence 1106

MEKRSINMKKVFMIIISILITITVTLACGGSGKQKEPSKESQKSDKYDYVYYEILNDGDSE
35 TPNVEIKYKDKKGKSHIEKADLDHVEHILGDGNKKPYIVKDGKKIHVYRPPYMIYGD
VEGKAVSKDEVTK*

Sequence 1107

Contig_0559_pos_2082_1474,
putative peptide of unknown function

atggtacttctcacgtcttctttaagtattgtcagtagcatattctcatgcaacaacgtca
40 ggaggaacgagtagttccagttcggcaagtcttagttcaagtagcagtgacgttctgca
tctagaggttcaacttcttcaagtacaagtatgagtcggttctagtgcaataaatgcgtct
cgcaatgcacaacaatctagtcagcgtgctgcccaacaagcaacaaaatcaagtcgtgta
acagcaacaaaaataaaggacaacaaagtgtatcaagacaaaaagcacaatctcggtct
45 ttgatgccgtctcaaagaccttatgattcaagtgaccatactcatctcaatatattgct
acaacttattataataattgggtattctattatatttttgacattcggtttttaaatcaa
catgaaaagaaaaacagtgtagatgctcagtttaatatgttgaaacaacaaatgaagcct
catgagaaactttatactgttactgtaaagactaaacaaggaaagcgtgctggttgtgta
cctaaaaacaatatgacaaaattgaaaaaggaaaacacattaaagttaaaaatggtggt
50 gttcagtaa

Sequence 1108

MVLLTSSLISIVSTYSHATTSGGTSSSSSSASSSSSSAASASRGSTSSSTSMRSSAINAS
55 RNAQQSSQRAAQQATKSSRVATKNKGQQSVSRQKAQSRSLMPSQRPYDSSAPYSSQYIA
TTYNNWLFYYIFAHSLNQHEKKNSVDAQFNMLKQMKPHEKLYTVTVKTKQGRVVVV
PKQYDKIEKGKHIKVKNGVVQ*

Sequence 1109

Contig_0561_pos_316_1254,

is similar to (with p-value 0.0e+00)
 >sp:sp|P45557|PRMA_STAAU_PROBABLE_METHYLTRANSFERASE (EC 2.1.1.-). >gp:gp|D30690|STANHS_5 Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds. NID: g4873

5 26.
 atgaattggatggaactctcaattgtagttaatcacgaagtagaatacgaatgttacagaa
 attcttgaaagttatggctctaattggagttgtaattgaagattcaaataattttagaagaa
 caacctattgataagtttgagaaatttatgacttaaacctgaagactatcctgaaaaa
 ggagttcgattaaaagcttactttaatgagttcacttataatgaaaacttaaaatccaac
 10 atcaattatgaaatattaagttcttcagcaaattgataaaaacaatttatgattaccaggaa
 aaacttattgcccgaagtagattgggaaaaatgaatggaagaattattttcatccatttaga
 gcttcaaaaacaatttacgatagtaccaagttgggaatcatatgttaaagaaaatgataac
 gaattgtgcattgaattagatccaggtatggcttttggaacaggtgatcatccaacgaca
 agtatgtgtttaaaagcaattgaaacttttgtaaaaccaactgattcagttatcgacgtt
 15 ggaacagggtcaggcattttaagtattgctagtcatttacttggagttcaaagaataaag
 gcattagatatagatgaaatggctgtaaatgtggcaaaaagaaaactttaagaaaaatcat
 tgtgatgatgcaattgaagcagttccaggttaatttattaaaaaatgaaaatgagaaattt
 aatatcgttattgcaaatattcttgctcatattattgaagaaatgattgaagatacttat
 aatactttaattgaagatggttattttatcacatcaggtattattgaagaaaagtatcaa
 20 gatatagaatcacaatgaagcgtattggtttcaaaattatttcagtagaacatgacaat
 ggctgggtttgtatagttgggtcagaaagtgaagtgataa

Sequence 1110

25 MNWMELSIVVNHEVEYDVTEILESYSNGVVIEDSNILEEQPIDKFGEIYDLNPEDYPEK
 GVRKAYFNEFTYNENLKSININYEILSLQQIDKTIYDYQEKLIAEVDWENENWKNYFHPFR
 ASKQFTIVPSWESYVKENDNELCIELDPGMAFGTGDHPTTSMCLKAIETFKPTDSVIDV
 GTSGILSIASHLLGVQRIKALDIDEMAVNVAKENFKKNHCDDAIEAVPGNLLKNENEKF
 NIVIANILAHIIIEEMIEDTYNTLIEDGYFITSGIIEEKYQDIESQMKRIGFKIISVEHDN
 GWVCIVGQKVSG*

Sequence 1111

30 Contig_0561_pos_1271_2065,
 is similar to (with p-value 3.0e-37)
 >sp:sp|P54461|YQEU_BACSU_HYPOTHETICAL_28.8_KD_PROTEIN_IN_DNA
 35 J-RPSU_INTEREGENIC_REGION. >gp:gp|D84432|BACJH642_115 Bacill
 us subtilis DNA, 283 Kb region containing skin element. NID:
 g2627063. >gp:gp|Z99117|BSUB0014_24 Bacillus subtilis compl
 ete genome (section 14 of 21): from 2599451 to 2812870. NID:
 g2634966. >gp:gp|D83717|D83717_3 Bacillus subtilis DNA for
 40 DnaJ, YqeT, YqeU, YqeV, YqeW, YqeX, YqeY, complete and parti
 al cds. NID: g1890057.

atgaatcaaagcgtgatgaaaatcagtgcttttttattgaaaacaaagaagactatcat
 catatcgatgaatgttatgctgctataaagaaggacaaaatattattgtcactttttcagat
 gaaaatgtattcaaatgtaaaattatttcaataaacgatcaatcgattgaaattaaatta
 45 gtagaaaagcaacaaattaacactgaactacctcagaacattacaatatgtagtggttta
 atcaaagcagacaaatatgaatggatgatacaaaaagcaactgaaatggggcaaatgag
 tttatagctgtagctatggaacgttctgtggtcaagctcaatgattctaaagtagaaaag
 aaattatcgagatggcaaaaaattataaaggaagctgcagaacaaagtattcggttaaca
 ataccaaatataaaatttaagtcgaatttaaaagaaatttatggtatgataagtcaatat
 50 gactatgttcttatagcatatgaagaacaagcaagcacggtgaatttaagtcatttaag
 caaacaattaaacaatttaagacacaggatcggtgttttaaatcatatttggaccaaatgaa
 gaaactaatgctgattctaaattcacaaaattttatcaaaaccaaactcgacaaactgaaa
 aatgcaaataacgctcaacttaataacgaaaatcaaagtaaaagttaacaacatgcttgaa
 gacatcaatacaaaaatttgatagatttaaagctaaactagaaaatatcttgatggatca
 55 aattcaggaaactaa

Sequence 1112

MNQSADENQCFFIENKEDYHHIVNVMRYKEGQNIIVTFSDENVFKCKIISINDQSIEIKL
 VEKQQINTELPQNITICSLIKADKYEWMIQKATEMGANEFIAMERSVVKLNDSKVEK

KLSRWQKIIKEAAEQSYRLTIPNIKFKSNLKEIYGMISQYDYVLIAYEEQAKHGELSQFK
QTIKQFKTQDRVLIIFGPNEETNADSKFTKFYQNQIDKLKNANNAQLNNENQSKVNNMLE
DINTKFDSIKAKLENILNGSNSGN*

5 Sequence 1113
Contig_0561_pos_7853_8494,
is similar to (with p-value 4.0e-52)
>gp:gp|299122|BSUB0019_48 Bacillus subtilis complete genome
(section 19 of 21): from 3597091 to 3809700. NID: g2636029.
10 >gp:gp|U56901|BSU56901_2 Bacillus subtilis putative transcrip-
tional regulator (yvhJ), Ycr59c/YigZ homolog (yvhK), histid-
ine kinase (degS), transcriptional regulator of degradation e-
nzyme (degU), (degV), (comFA), (comFB), (comFC), flagellar p-
rotein (yviB), negative regulator of flagellin (flgM), flage-
15 llar protein (yviC), flagellar-hook associated protein 1 (flgK),
flagellar-hook associated protein 3 (flgL); (yviE), transmembrane
protein (yviF), (csrA), flagellin (hag), flagella-
r protein (yviH), flagellar hook-associated protein 2 (fliD),
flagellar protein (fliS), flagellar protein (fliT), sigma-
20 54 modulator homolog (yviI), and (secA) genes, complete cds.
NID: g1762326.
atggataaatccataattactattaaacaagcacattcaattgaaaatgtgataagtaaa
tcacgcgtttatagcatatattaagcctgtttcgactgaaaatgaagcaaaagctttata
gatgaaattaaaacaaaacataaagatgcaactcataattgttcagcctatactgtcgga
25 ccagagatgaatattcaaaaggcaaacgacgatggcgaaccaagtggacacgtggcatc
ccaattgcttgaaatactgaaaaacaagagatacacaatgtttgtgtcgtcgtgacacgc
tacttcggtggtatcaagtttaggtgcagcggtcttattagagcatatagcggcgccgtg
cgtgatgtgatatatgatataggtagagtcgaactaagagaagctattccagtaaccggt
acggttagattatgatcagacaggtaaatttgaatatgaacttgccctctactacattctta
30 ttaagagaacaattttataccgataaagtaagttatcaaattgacgtagtaaaaaatgaa
tatgatgcttttatagacttttttaaatcgaattacttctggaattatgatttgaaacaa
gaagaccttaaaactattaccttttgatattgaaaccaattaa

Sequence 1114
35 MDKSIITIKQAHSIENVISKSRFIAYIKPVSTENEAKAFIDEIKTKHKDATHNCSAYTVG
PEMNIQKANDDGEPSGTAGIPMLEILKKQEIHNVCVVVTRYFGGIKLGAGGLIRAYSGAV
RDVIYDIGRVELREAIPTVTVLDYDQTGKFEYELASTTFLLEQFYTDKVSQIDVVKNE
YDAFIDFLNRTSGNYDLKQEDLKLPLFDIETN*

40 Sequence 1115
Contig_0561_pos_9667_8555,
is similar to (with p-value 0.0e+00)
>pir:pir|A55856|A55856 llm protein - Staphylococcus aureus
gtgaggatcaacttattcaatgaaggatgaactgatgtataactattactttagctttt
45 actatgatagtcagtttaataattacacccattattattgtaatatcaaaaaaattagat
ttagtagatcgtcctaatttcagaaaagtacatacgaaacctatctcagtgatgggagga
acggtcatttttattttcttcttaattagggatttggctcggacaccctattgaacgtgag
gttaaacgcgttatattaggtgcaattacaatgtatatggttggtgattgattgatgatt
tacgatctaagaccttatttaaagtttagcaggtcaaattgttgacgctttaattgttacg
50 ttttatggaattacaatagactttatttccattgccaattggtccaacgattcattttggc
atattcagcattcctattacagtaatatggattgtagcaattaccaatgctattaatctt
atcgacggacttgatggacttgccctcaggcgtctcagcattggcattaatgactattgga
ttcatcgctattttacaagcgaacataatttattatcatgatttgcgtgtacttttaggg
tctttacttgggtttcttattctataacttccaccagcgaataatttcttaggtgatagt
55 ggtgcattaatgataggattttattatcggtttcttactcctactcggttttaagaatatac
acattttattgcattattcttctctatagttatattagcgggtgccattttattgatacatta
tttgcaatgatttcgtcgaatgaaaaaagggaacatatataatgcaagcggacaagtcacat
ttacatcataaattacttgccttaggatatacgcatagacaaaccgttttacttatttat
tcaatagcgattatgttttagtttatctagtgttatcctctattttatcccaaccgttgggt

gcacttatgatgttcattctcattgtctttacgattgagttgatcggtgaatttactgga
ttaatagatgataattatcgaccaatattaaatttaattacaaaaaaggaaatggtaag
caacatcattatgatgagcatcaccgttcataa

5 Sequence 1116

VRYNLFNEGELMYTLLLIAFTMIVSLIITPIIIVISKKLDLVDNPNFRKVHTKPIISMGG
TVILFSFLIGIWLGHPIEREVKPLILGAITMYMVGLIIDDIYDLRPLYKLAGQIVAALIVT
FYGITIDFISLPIGPTIHFGIFSIPITVIWIVAITNAINLIDGLDGLASGVSALALMTIG
FIAILQANIFIIMICCVLLGSLGFLFYNFHPAKIFLGDSGALMIGFIIIGFLSLLGFKN
10 TFIALFFPPIVILAVPFIDTLFAMIRRMKKGQHIMQADKSHLHHKLLALGYTHRQTVLLIY
SIAIMFSLSSVILYLSQPLGALMMFIVFTIELIVEFTGLIDDNYRPILNLITKKGNKG
QHYYDEHHRS*

Sequence 1117

15 Contig_0561_pos_7708_6842,
is similar to (with p-value 1.0e-48)
>sp:sp|P32436|DEGV_BACSU DEGV PROTEIN. >pir:pir|S28596|D3019
1 hypothetical protein U3 - Bacillus subtilis >gp:gp|Z18629|
BSCOMFG_1 B.subtilis comF gene. NID: g39847. >gp:gp|Z99122|B
20 SUB0019_45 Bacillus subtilis complete genome (section 19 of
21): from 3597091 to 3809700. NID: g2636029. >gp:gp|U56901|B
SU56901_5 Bacillus subtilis putative transcriptional regulat
or (yvhJ), Ycr59c/YigZ homolog (yvhK), histidine kinase (deg
S), transcriptional regulator of degradation enzyme (degU), (degV),
25 (comFA), (comFB), (comFC), flagellar protein (yviB),
negative regulator of flagellin (flgM), flagellar protein (yviC),
flagellar-hook associated protein 1 (flgK), flagellar-hook
associated protein 3 (flgL), (yviE), transmembrane prot
ein (yviF), (csrA), flagellin (hag), flagellar protein (yviH
30), flagellar hook-associated protein 2 (fliD), flagellar pro
tein (fliS), flagellar protein (fliT), sigma-54 modulator ho
molog (yviI), and (secA) genes, complete cds. NID: g1762326.
atgaagattgcagttatgaccgattctacaagttatttaccacaacatataatagaacaa
tataacataccagtcgcttcactaagtgtactttcgatgatggagtgaatttcactgag
35 agttagatgatttttctgtagatgatttttataaaaaaatggcttcactctaaactatacca
acaacaagccaacctgctattggcgattggattgaaaattttgagagattaagagaaacaa
ggatacactgatgtcatcggtgattaacttatcaagtggtataagcgggaagctatccttca
gcaacacaagctgggtgaaatggttgaagatattcaagtagacatacggttgatagccgtcctt
gctgcagatgattgaggttagctttgcaatttacgctgctcaattgggtacaaaaggat
40 aaacctgatgatattattaatgaactaactgaaataagacaacatattggtgcatactta
attggtgatgatttaaaaaattacaaaaaagtggtcgtatcactggagctcaagcttgg
gtaggtacattattgaaaatgaaacctgtcttgcgttttgaagaagatggtaaaatacat
ccacacgaaaaagtagctactaaaaaacgtgcgctaaaatctttagaaacaaacattttt
aaagaaatagaaggcatggaagatgtgacagttattgtataaacgggtgataaaactgaa
45 gatggaaagtcattttctcagcaattaaaggaagatcatcctaattgttcataattcagtat
tgtgaatttggaccagtgatagcatcacatttaggatcaggcggtttaggattgggttac
ttcccaagaagaatcgacattaattaa

Sequence 1118

50 MKIAVMTDSTSYLPQHIIIEQYNIPVASLSVTFDDGVNFTESDDFSVDDFYKKMASSKTIP
TTSQPAIGDWIENFERLREQGYTDVIVINLSSGISGSYPSATQAGEMVEDIQVHTFDSRL
AAMIEGSFAIYAAQLVQKGYKPDIIINELTEIRQHIGAYLIVDDLKLNQKSGRITGAQAW
VGTLLKMKPVLRFEEDGKIHPHEKVRTRKRALKSLETNIFKEIEGMEDVTVFVINGDKTE
DGKSFLQQLKEDHPNVHIQYCEFGPVIASHLGSGLGLGYFPRRIDIN*

55

Sequence 1119

Contig_0561_pos_6305_5436,
is similar to (with p-value 4.0e-44)
>sp:sp|P39145|CMF1_BACSU COMF OPERON PROTEIN 1. >pir:pir|S28

597|S28597 hypothetical protein Fl - Bacillus subtilis >gp:gp
 p|Z18629|BSCOMFG_2 B.subtilis comF gene. NID: g39847. >gp:gp
 |Z99122|BSUB0019_44 Bacillus subtilis complete genome (secti
 on 19 of 21): from 3597091 to 3809700. NID: g2636029. >gp:gp
 5 |U56901|BSU56901_6 Bacillus subtilis putative transcriptiona
 l regulator (yvhJ), Ycr59c/YigZ homolog (yvhK), histidine ki
 nase (degS), transcriptional regulator of degradation enzyme
 (degU), (degV), (comFA), (comFB), (comFC), flagellar protein
 (yviB), negative regulator of flagellin (flgM), flagellar p
 10 rotein (yviC), flagellar-hook associated protein 1 (flgK), f
 lagellar-hook associated protein 3 (flgL), (yviE), transmembr
 ane protein (yviF), (csrA), flagellin (hag), flagellar prot
 ein (yviH), flagellar hook-associated protein 2 (fliD), flag
 ellar protein (fliS), flagellar protein (fliT), sigma-54 mod
 15 ulator homolog (yviI), and (secA) genes, complete cds. NID:
 g1762326.
 atgggacataatattgctattgtatcacctcgtgtagacgttattattgagataagtc
 cgaattaaagatgcttttatcgatgaacatatagatgtgctacatcaatctagtagacag
 caataaatggctcattttgtattgctactatccatcaattattgaggtttaaacagcat
 20 tttgatactgtattttgtcgatgaggttagatgcttttccgttgctctatggatccacaatta
 tcaaatgcaatacaacttgcttcaaaatcgaatcattcacataattttcatgacgcccaca
 ccaccgctgctatttttaaaacaattccccccagaaaaataattaagttaccagcccg
 tttcaccgcatcccccttccattcctaagttcaaatatttcaaattaaaatcaacacga
 aaacaaaattttattacttaatatatttagatatcaaatcaacacacgcttttactttg
 25 gtctttattaataatatagaattatgaataaaatgtatcaacagtataaaatggacatc
 cctgattgtattgcttcacagtgaaagatgatttacgatttgaaaaaattgaagcttta
 agacgaggacaacacaaaattgtattcactacaactattttagaagaggatttacaatg
 acacacttagatgtcgttgtagttgatgctggaagttttcaacaagaggctttaattcaa
 attgctggctgcgtaggacgtaaacagcagctctccaagtggttagttttattttctcat
 30 gaaggtgttacattatcgatgatttttagctaaaagaacattatttcaatgaatcggtta
 gcaattaaaaggggatggattgatgcgtaa

Sequence 1120

MGHNIAIVSPRDVIIIEISHRIKDAFIDEHIDVLHQSSRQQYNHGFVIATIHQLLRFKQH
 35 FDTVFVDEVDAPFLSMDPQLSNAIQLASKSNHSHIFMTATPPRHFLKQFPPEKIIKLPAR
 FHRSPLEPIPKFYKFLKSTRKQNLNIFRYQINQQRFTLVFINNIEIMNKMYQQYKMDI
 PDLICVHSEDDLRFKIEALRRGQHKIVFTTILERGFTMTHLDDVVVDAGSFQQEALIQ
 IAGRVGRKQQSPSGLVFLHEGVTLMSILAKRNIISMNRLAIKRGWIDA*

40 Sequence 1121

Contig_0561_pos_5227_4769,

putative peptide of unknown function

atggaacaattgttttgtgattatagttatgatggatgatgaaagaaatcatcacaccag
 tataagattaagcgagacttctatttggcagaagtattggcgagaaaattagttttacct
 45 caaacgcaatatgattatatagttccattccttctccaattgaacgcgacattgaacgt
 acatttaactcctgtgaccactgtcttagataaaatgggcattctcatatcaagatgtatta
 ggtacacatatatcgtcctaagcagtcctaaagtaggaagattgaacgttcaaaagccct
 aatccattttatataaaaagatgaagagataaatatcgaagggaagtaataactactcata
 gatgatatttatacaacagggttaactattcatcacgcagggtgtaaattgtacgataaa
 50 aaagtcagaaaattcaaaagtgtttgcgtttgcacgataa

Sequence 1122

MEQLFCDYSYDGMMEKIIHQYKIKRDFYLAEVLARKLVLPQTQYDYIVPIPSPIERDIER
 TFNPVTTVLDMGISYQDVLGTHIRPKQSKLGKIERKAPNPFYIKDEEINIEGKVILLI
 55 DDIYTTGLTIHHAGCKLYDKKVRKFKVFAFAR*

Sequence 1123

Contig_0562_pos_5078_5830,

putative peptide of unknown function

atgaatgccatgaaagataataactaataaattgcatcaagcggttaactaaaatacaacaa
 aaaatgcccggggatggggagacacgcctcatcaagatatggctaaaccatataaacta
 acaacgcactatattatattggttcacagattctacgtatattgatatgataaatcctatt
 ttaattggatattttgtctttttctttacgtttttaatttctggcattggcttattaaaa
 5 gagcgtacttctggcacattagaacgtttacttgcctctccaataaaaagaagtgaatt
 attttgggttatgttttcggttatggtagtttttagcggtatccaaacaatagtgtcgta
 ttatatgcaatttatattctgcatatagacttagtaggttcgatatggttcgtactatta
 acggcaatattaacagcgcttctgctgtgacattcggtatattattatctacctttgct
 10 tcttcagaattccaaatgattcaatttataccatttagtcatagtccacaagtactattt
 gcaggcattataccaattgaatcaatgaataaaggattacaatactttcacatatcatg
 ccgttattctataccggccaaacgatgcaaaatattatgatcaagggttatggattcaac
 gatatttacatttatttaattgtgttatttcgcatttttctatttctattgattttaaat
 attataggcatgaaaagatatagaaaagttag

15 Sequence 1124

MNAMKDNNTNKLHQALTKIQKMPGDGGDTPHQDMAKPYKLTTHYLYGSSDSTYFDMINPI
 LIGFFVFFFTFLISGIGLLKERTSGTLERLLASPIKRSEIIFGYVFGYGSFSVIQTIWV
 LYAIYILHIDLVGSIWFLVLLTAILTALVAVTFGILLSTFASSEFQMIQFIPLVIVPQVLF
 AGIIPIESMNKGLQYFSHIMPLFYTGQTMQNMIMIKGYGFNDIYIYLIVLFAFFIFLLILN
 20 IIGMKRYYRKV*

Sequence 1125

Contig_0562_pos_5840_6490,
 putative peptide of unknown function

25 atgaaccaagatatattaagtcatttagttgaaaccattgtgcctcaacttgaatatattaagc
 gataaacaagacgtgtcatagaaagtgcatttgcattattcagtgaacaaggatttgat
 aaaacgaagtactaaagaaattgcgcagcgtgcaaatgtcgcagaaggaacggtatttaag
 cagtttaaaagtaaaagaatgttattatacgcaggattaattccaattttaagagatcat
 atcgcacctgtagctgttaacaatttacagatgaattaaacgaagtaaccattttgat
 30 gcatttataaattttattgtagaaaatagatctaaattttatttatgacaatagacgtatt
 cttaaagtcattcttaaatgaagctattactaatgaagattttcaaaatatattagttaat
 attttccccataaattaacgagtaaaattaaaagataaaaattgaatgggtttatcgataat
 ggtgacatgcgcaatgttaaacctgagtttttatcgtacggtcgctgcacaaatttta
 aatttaaatatcccaataatagtttaataatgactataactaagggtgaaaactatcagcag
 35 ttgtcggttattcgtataaaagagggttatataggatgtttaagcgagaatag

Sequence 1126

MNQDIKSLVETIVPQLEYLSDKQRRVIESAIALFSEQGFDKTSTKEIAQRANVAEGTVFK
 QFKSKRMLLYAGLIPILRDHIAPVAVKQFTDELNEVTHFDAFINLFVENRSKFIYDNRRI
 40 LKVILNEAITNEDFQNILVNIETHKLTSKLDKIEWFIDNGDMRNVKPEFFIRTVVAQIL
 NLNIPPIVNNDYTKGENYQQFALFVKEGLYRMFKRE*

Sequence 1127

Contig_0562_pos_7987_8625,
 putative peptide of unknown function

45 atgaaattagaggaaattaatacaatagataaaaatgatttttctaaaaaaccaacttat
 ggggctgaaaaaaattattttaaaattgctgttatggattttcaaaaagagagtacaaag
 agactagacattgatatatatgatactggaactaagatagcaaaatattggagcgaagga
 tttttatcaatttatccgttgagagataatgaattaaataactaatgatattattaactct
 50 ttaaaagaaaatactctgttcaatgaattacctaataacagaaaaagatcttaatacaa
 tctactaatgactttctaaagtataatagagaattttcatttgtaaaattaagagatcat
 attaatcagcgctttgaaacagacttcaaaaatgatgaactatttaagatgacaaatta
 aatcatatagacgcagaatttaatttagaaccaaggattgtaggttaaatatttacagaaa
 55 aaaataaaaattaatgatgtaatcagtttagatattaataatattagtaaaagcaacaaa
 ggaggcctaataaagtgcttaataataaaagataacttaaaaatacgtctcgaaacaaat
 attaaagaagaccttcccgatattcaggaaggagattga

Sequence 1128

MKLEEINTIDKNDFSKKPTYGAENYFKIAVMDFAQKESTKRLDIDIYDTGTAKIYWSEG

FLSIYPLRDNELNTNDIINSLKENTLFNELPNITEKDLNQSTNDFLSDNREFSFVKLRDH
INQRFETDFKNDLFDKDDKLNHIDAEFNLEPRIVGKYLQKKIKINDVLSLDINNISKATK
GGLMKVSNDRYLKIRLETNIKEDLPDIQEGD*

5 Sequence 1129

Contig_0562_pos_8629_9843,

putative peptide of unknown function

atgagttattattggtattatgaaatTTTTAAATTTGATAATGATTTAAATAATTATTTT
tctaataagacttaaaagtTTTTTCTTCTAAATGTCTTCAATGAAGAAGAAGATGAA
10 ttatatttacaaggatcaaaaatatttattgatgacgactctaaaattaaacattcagaa
agtattaaagatgtgtatgagtactattatgaaaaattcaggaaattagaaaattctaaa
tctgattTTTTTATATTAAAGACAGTAAAAAGTTATTTGAAAAATTGAAGTACCTATT
aaacataatacgcatacgcaggagaatatattgaagataattactgtcagaaatataaatagc
aatTTTTTGCAAAATAATCCTTATTTTTCAATGAAATGCTGGAAGTATATAAAGATAT
15 gaaaaagaatgcagtaataataaggaagttgcgtaattacaatctgcaatcattttatct
gataaaatttcatttgataacgcataatcatgacattatactttcatgattcagataatgaa
gaagataagttccaaaagaacaagtttgattttcaaacgcataatgtatcttcattatat
tttatgttttttgattgggtagatattgaaaacgcataatctagaaatagtggatttaaaa
tataaagtagcaaaaattttatttctaaattcgaaaagtttctgatgacgaaaaaata
20 aagaaaaaagaatacatcatgacttaaatgttatgtataatttaatacttcaaaaaaaa
tctcaaaagtattatgaatataacaaagtaataaggaatcataaaatagaaatcattcaa
cgaaaaatagaattaaaaaatgaacttaataaaaaaattaatgagcatgatggtattcatt
cccgtaactatttatggtttatatataacaatccaaaagagcaagaatcactaaacatt
ttcaataatgactttaataattattttcagttctctagttgcacttatatttataata
25 ttatcttttaataatgatgtgaaatctattaatagtgactatgaaacaattattttagaa
atcataaatacatataaaataaacaagaatggacgattttggaaataatataagttta
agtgatttttaattttctttattttggattttttataatgattataactcttataattttt
attttaattaagtga

30 Sequence 1130

MSYYWYIEILKFDNDLNNYFSNEDLKVFSKCLSNEEDELYLQGSKIFIDDDSKIHKHSE
SIKDVYEEYIEKFRKLENSKSDFFIFKDSKKLFKIEVPIKHNTIRGEYIEDITVRNINS
NILQINPYFFNEMLEVYKEYEKECSNIRKLRLNLSAIIILSDKISFDNDIMTLFHDSDNE
EDKFQKNKFDNANVSSLYFYVFDWVDIENDNLEIVDLKYKVAKIYLSKFEKVSDEKI
35 KKKEIHHDNLVMYNIILQKKSQKYIEYNKVIRNHKIEIIQRKIELKNELNKKLMSMMVFI
PVTIYGLYITIQKSKESLNIFNDFNIIFSSSLVALIFIILSLINDVKSINSDYETIILE
IINTYKINKSMDDFGNNISLSDFKFSLFWIFIMIITLIIFILIK*

Sequence 1131

40 Contig_0562_pos_13601_13936,

is similar to (with p-value 3.0e-38)

>gp:gp|AF051916|AF051916_2 Staphylococcus aureus plasmid pJE
1 remnant of replication protein Rep (rep), trimethoprim res
istance protein DfrA (dfrA), thymidylate synthetase ThyE (th
45 yE), and putative transposase Tnp (tnp) genes, complete cds;
and unknown gene. NID: g3676404.

atgaatgatataactaaacgtttatttaaaaccaataattaatgagctttcttcaattttt
aataaccttcatattaataagatcaaagctaaaaaggacgtaaaattgaatggtagag
tttacctttgacgctgagaaacgcattcacacaagcgacaaccacaaatgactaatata
50 ggtaagtcgcgccaatataccaatcgtagaaaaacacctaataatggtagacgaaaagata
tataaacaatctcaagagatacataatgaagatgcaaaattaaaacaagatcgagaggca
tttcaacgtcaattagaagaaaaatgggaggaataa

Sequence 1132

55 MNDITKRLKPIINELSSIFNNLHINKIKAKKGRKIEWLEFTDAEKRIHNKRQPMQMTNI
GKSRQYTNREKTPKWLDKIKYQSQEIHNEDAKLKQDREAFQRQLEEKWEE*

Sequence 1133

Contig_0562_pos_7830_7156,

is similar to (with p-value 0.0e+00)
 >pir:pir|S04166|S04166 transposase 2 - Staphylococcus aureus
 plasmid pSK1 transposon Tn4003 >gp:gp|X13290|SATN4003_5 Sta
 phylococcus aureus multi-resistance plasmid pSK1 DNA contain
 ing transposon Tn4003. NID: g46747. >gp:gp|U40259|SEU40259_7
 5 Staphylococcus epidermidis trimethoprim resistance plasmid p
 SK639. NID: g1762079. >gp:gp|U40382|SEU40382_1 Staphylococcus
 epidermidis plasmid pSK697 insertion sequence IS257(697B) p
 utative transposase gene, complete cds. NID: g1762093. >gp:g
 10 p|U40383|SEU40383_1 Staphylococcus epidermidis plasmid pSK697
 insertion sequence IS257(697C) putative transposase gene, c
 omplete cds. NID: g1762095. >gp:gp|AF051916|AF051916_3 Staph
 ylococcus aureus plasmid pJE1 remnant of replication protein
 Rep (rep), trimethoprim resistance protein DfrA (dfrA), thy
 15 midylate synthetase ThyE (thyE), and putative transposase Tn
 p (tnp) genes, complete cds; and unknown gene. NID: g3676404
 . >gp:gp|AF051917|AF051917_15 Staphylococcus aureus plasmid
 pSK41, complete sequence. NID: g3676412. >gp:gp|AF051917|AF0
 51917_17 Staphylococcus aureus plasmid pSK41, complete seque
 20 nce. NID: g3676412.
 atgaactatttcagatataaacaatttaacaaggatgttatcactgtagccgttggtac
 tatctaagatatgcattgagttatcgatgatataatctgaaatattaaggggacgtggtgta
 aacgttcatcattcaacggtctaccgttggttcaagaatatgcccccaattttatatcaa
 atttggaaagaaaaagcataaaaaagcttattacaaatggcgtattgatgagacgtacac
 25 aaaataaaaggaaaatggagctattttatctgctgccattgatgcagaggacatacatta
 gatatttgggttcgctaagcaacgagataatcattcagcatatgcgttttatcaaacgtctc
 attaacaatttggtaaacctcaaaaggtaattacagatcaggcaccttcaacgaaggta
 gcaatggttaaagtaattaaagcttttaaaccttaaacctgactgccattgtacatcgaaa
 tatctgaataaacctcattgagcaagatcaccgtcatattaaagtaagaaagacaaggat
 30 caaagtatcaatacagcaaaagaataacttttaaagggtattgagtgatattacgctctatat
 aaaaagaaccgcaggtctcttcagatctacggattttcgccatgccacgaaattagcatc
 atgctagcaagtttaa

 Sequence 1134
 35 MNYFRYKQFNKDVITVAVGYLLRYALSYRDISEILRGRGVNVHSTVYRWVQEYAPILYQ
 IWKKKKKKAYYKWRIDETYIKIKGWSYLYRAIDAEGHTLDIWLKQQRDNHSAFYKRL
 IKQFGPKQKVITDQAPSTKVAMVKVIKAFKLPDCHCTSKYLNLIQDHRHIKVRKTRY
 QSINTAKNTLKGIECIYALYKKNRRSLQIYGFSPCHEISIMLAS*

 40 Sequence 1135
 Contig_0562_pos_4253_3876,
 putative peptide of unknown function
 atgcaattttattatagtaattgtgaacagaatgcatcaaaacttatttagtcatcaacgta
 caacctaataagaggattttctttatgtgtgaatggtaagaaaagtaatacaaaataatgaa
 45 atgcaaaaagtgaagctttcttatactatgccgattaaagataaaatgaacacagttgat
 gcatatgaaaaatcttatttacgatacattaattggagaacaaacaaaatttacgcattgg
 gaagaattaaaaattcttggaaattttattgatgatattgaaaatgtatggaaacaagaat
 agccacagtttcttaattatgcctttggatgctatgggcctaaagaaagtgaaaaattac
 ttagtgaagacggattga
 50
 Sequence 1136
 MQFYYSNCEQNASNLLVINVPNEGFSLCVNGKKSNNEMQKVKLSYTMPKDKMNTVD
 AYENLIYDTLIGEQTFTHWEELKILGNLLMILKMYGNKNSHSFLIMPLDAMGLKKVKNY
 LVKTD*
 55
 Sequence 1137
 Contig_0562_pos_3757_1499,
 is similar to (with p-value 0.0e+00)
 >gp:gp|L76359|STMDRRC_1 Streptomyces peucetius daunorubicin

resistance protein (drrC) gene, complete cds. NID: g1196906.
 atggatttttattaatattacaggtgcttcacaaaataacttgaaaaacatagatgtaaat
 atcccaaaacacttagtaacgggtatttacaggtcggttcagggaatcatcttta
 gtgtttaatactgttgctgcggaagtctgaacagctactaaatgaaagttattctagttat
 attcaatttcattttaaatcaacaacccagaccgaaagttaaagaaaattaaaaatcttcct
 5 gttagcaatgacgattaatcagaaaagattcaatgggaattctcgctccacggtaggaaca
 gtttcagatatatatgcttctgttagattactgtggtctagaataggcgaaccgtttgtt
 ggttattcagatgcatattccttcaatagtcctaaggcgatgtgtaaaacttgtgagggga
 ttaggatataattgaagacattaacttagatgaattgctagattgggataagtctttaaat
 10 gaaggtgcaatagactttccttcttttgaccagacaaagagcgtggtaaagcctatcga
 gatagtgggtttatttgataataataaaaaattgaaagattatacagaagatgaattagaa
 ttgtttttatatcaagagccaatgacattaaaaaatcctcctaagaatggagaaagtca
 gctaaatatgtaggactaatacctagattcagtagaataatttttaggtgataaagaattt
 aataagaaacgctacgccaacatcttaaaaaatgtagtaataataaaaaatctgttcaaca
 15 tgtaaaggtcaacgtctaaactcgaaaatattaagttctaaaattatgagtaaaaaatatt
 tctgatttcacacaaatgacaattaaggaaaatttagagtttcttaataaatttagaggat
 ccaacagccaaatatattattgatcctctcaaaaagcagttagaagcactagaatatatt
 ggattaaagttatttaacgcttaaccgtgtcacaacgacattatcaggcgggtgaagcgcaa
 cggcttaaatattaacgtcatttaaatagttctttatcggatttagtttacattatagat
 20 gaaccaagtgttggttgcacccggaagatatagctaaaaatcaatgaaattttaaaatca
 ttaaaagaaaaaggtataactgtgttaattgttgaaacatgatcccgatgtcattaaagaa
 ggagactatatcatagatatggggccaggttcaggaaaaaacggcgggtgaaatcacattt
 gaaggaacatataatgaattactatctcaaaatacttcgacaggttaacgcattacgtaac
 aaacataattttaaaagagaatattcgtgaagctaaccactttataatatcggctcctgtg
 25 acacaaaacattttaataacgtaaaaacgtctatacctaaacacgtattaaacagtcctta
 acaggtgttgctggttcaggtaagagtaaccctgtttaaagcaggttttgaaaataatgac
 cataccatctttattgatcaaaaagcagttcaaggatctaatagatctaatactattaacg
 tatttaggtgttttgatagtgtaagaagctatttcagtaagaaacaggcttaataaaa
 gctatgttttagttataattcaaaaggtgcctgtccaaattgtggtggaaagggctatatt
 30 aaaacggaacttgcttttatgggtgatttttcacagacatgtgaagtttgcacatggcaaa
 cgttataaacaagaagtagtagatgctaccatagacgggtattcaattgccgatgttctc
 aatttgacgggttgacgaaggtatcattttctttgataaaaaagaatgatattaagtcaaaa
 ttacaatctgttaagtaagacaggtttgaattatatgtcactaggacaacctttgtccact
 ttgtctggtggagagattcaacgcgtgaaactaggacaacatttagatgaagagataaag
 35 aatagtagttttttatttttgacgaaccaactacaggcctacatgaatcggatatccctata
 ttgatggagtggtttgatgatttaattgatcaaaaacaatactgttatttgattgaacat
 aatttatcgattatgtgtgaagcagattggatcatcgatgtcggccagggccagggttg
 gatggcggaaaggtccaatttagtggaacacctaataaacttcattgatagttcagaaaca
 ttgacatctaaccacttgaaacgctatatcaaacagtaa

Sequence 1138
 MDFINITGASQNNLKNIDVNI PKHLVTFTGRSGSGKSSLVFNTVAAESEQLLNESYSSY
 IQFHLNQPRPKVKKIKNLPVAMTINQKRFNGNSRSTVGTVSDIYASVRLLSRIGEPFV
 GYSDAYSFNSPKGMCKTCEGLGYIEDINLDELDDWKSINEGAI DFPSFGPDKER GKAYR
 45 DSGLFDDNNKKLDYTEDELEFLYQEPMTLKNPPKEWRKSAKYVGLIPRFSRIFLGDKEF
 NKKRYAKHLKNVNNKICSTCKGQRLNSKILSSKIMSKNISDFTQMTIKENLEFLNKLED
 PTAKYIIDPLKKQLEALEYIGLSYLTNLRVTTTSLSGGEAQRLLKIRHLNSSLSDLVYIID
 EPSVGLHPEDIAKINEILKSLKEKGN TVLIVEHDPDVIKEGDYIIDMGPGSGKNGGEITF
 EGTYNELLSSNTSTGNALRNKHNLENIREANH FYNIGPVTQNNLNNVKTSIPKHVLTVL
 50 TGVAGSGKSTLVKAGFENNDHTIFIDQKAVQGSNRNLLTYLGVFDSVRSYFSKETGLNK
 AMFSYNSKGACPNCGGKG YIKTELAFMGDFSQTCEVCHGKRYKQEVLDATIDGYSIADVL
 NLTVDEGIIFFDKKNIDIKSKLQSVSKTGLNYMSLGQPLSTLSGGEIQRVKLGQHLDEEIK
 NSIFIFDEPTTGLHESDIPILMECFDDLIDQNNVTILIEHNLSIMCEADWIIDVGP GPGL
 DGGKVQFSGTPKNFIDSSETLT SKHLKRYIKQ*

Sequence 1139
 Contig_0563_pos_4995_4084,
 putative peptide of unknown function
 gtgattaggaatttacttgcaatgtgttattttgtatctaggtgagtagatgatagcgccaaa

gcaatgtttgaagaacttttaaggaagataattcagacgtgcatgcactttgtcactac
acattattactttataataaaaaagaaacagaaaaatatcaaaaatatcttaaaatactt
aataaagtagtaccactaaatgacgacgaaaccttttaaattaggaatcgtattgagttat
5 ttaaaacagtatcgtgcttctcaaaatttactttatccactttataaaaaaggtaaattt
gtctctattcaaatgtataatgcattgagtttcaattttttattacctaggaaataaagac
gaaagtattgagatgtggaacaagctcactcaaatttctgaagttgatgttggttatgca
ccttgggtaattgaggaaagtaaaacggtatttgaatcacgagtggtaccattattacta
gatgataataatcattatcgactttacggtatttttttacttcatcaattaaatggaaaa
10 gaaataactaatgactgaagatatttgggtcaattcttgaatcaatgaatgactatgagaaa
ctttatctcacatatttgggtacaaggactcacactcaataaattagattttatacacaga
ggatgcaaagggtgtataattttaagaaattcaaatataaacacgctctttatttacagat
tggattaatcaagcagaaatgattatagctgaaaatgtagatttagtagatgtcgataga
tatgtagctgcatttgtttacctatcgatcgctgcttctagccaaccacttaccaaggag
caattgatggacgatttttaattgtttctagatacaaaactgaataaagcaattgaatttata
15 ttgagcatataa

Sequence 1140

VIRNLLAMCYLYLGEYDSAKAMFEELLKEDNSDVHALCHYTLTTYLNKKETEKYQKYLKIL
NKVPLNDDETFKLGIVLSYLKQYRASQNLPLYKKGKFVSIQMYNALSFNFYILGNKD
20 ESIEMWNKLTQISEVDVGYAPWVIEESKTVFESRVLPLLLDDNNHYRLYGIFLLHQLNGK
EILMTEDIWSILESMDYKLYLTYLVQGLTLNKLDFIHRGMQRLYNFKKFKYNTSLFTD
WINQAEMIIAENVDLVDVDRYVAAFVYLSYRRSSQPLTKRQLMDDFNVSRYLKLNKAIEFI
LSI*

25 Sequence 1141

Contig_0563_pos_4019_3087,

is similar to (with p-value 0.0e+00)

>gp:gp|AJ223781|SAAJ3781_1 Staphylococcus aureus trxB gene.

NID: g3582102.

30 atgactgaagtagattttgatgtagcaataatcggtgcaggtcctgccggtatgacagca
gcagtatatgcatctcgtgccaatttaaaaactgtcatgattgaacgcggtatgccaggc
ggcctaaatggcaaacactgaagaagtagagaattttccaggatttgagatgatcacaggt
cctgacttatctactaaaatgtttgaacatgctaaaaaatttgggtgcggaataccaatat
ggcgatattaaatctgttgaagataaaggcgactataaaagttatcaatttagggaataaa
35 gagataacagcacatgcagttattatctcaactggagcagagtataaaaagattgscgtt
cctggtgaacaagaattaggaggacgtggagtaagttattgtgagggttggatggagca
ttctttaaaaataaacgtcttttcgtaattggcgcgagattcagcggtagaagaaggt
actttcttaactaaatttgcagataaagtaacgattgttcaccgtagagatgaattacgt
gcacaaaacacttgcagaacgtgccttcaaaaatgataaagttgactttatttggagt
40 ctacactttaaacaattaatgaaaaagatggtaaaagtgggtcagttacacttgaatca
actaaagatgggtgctgaacagacttatgatgccgacggtgtattcattttatattggaatg
aaaccactcacagcaccatttaaaaatcttggattacaaatgacgcgggatacattgtc
acacaagatgacatgagtactaaagtacagggtattttgctgcaggtgacgttcgtgat
aaagggttacgtcaaattgttactgctacaggagacggtagattgaggctcaaagtga
45 gctgattatattacagaattaaaagataattaa

Sequence 1142

MTEVDFDVAIIAGAPAGMTAAVYASRANLKTVMIERGMPGGQMANTEEVENFPGFEMITG
PDLSTKMFHAKKFGAEYQYGDIKSVEDKGDYKVINLGNKEITAHAVIISTGAEYKKIGV
50 PGEQELGGRGVSYCAVCDGAFFKNKRLFVIGGGDSAVEEGTFLTKFADKVTIVHRRDEL
AQNILQERAFKNDKVDFIWSHTLKTINEKDGKGVSVTLESTKDGAEQTYDADGVFIYIGM
KPLTAPFKNLGITNDAGYIVTQDDMSTKVRGIFAAGDVRDKGLRQIVTATGDGSIAAQA
ADYITELKDN*

55 Sequence 1143

Contig_0563_pos_2911_2003,

is similar to (with p-value 3.0e-88)

>sp:sp|O06973|YVCJ_BACSU HYPOTHETICAL 33.9 KD PROTEIN IN CRH
-TRXB INTERGENIC REGION. >gp:gp|Z99121|BSUB0018_163 Bacillus

subtilis complete genome (section 18 of 21): from 3399551 to 3609060. NID: g2635827. >gp:gp|294043|BSZ94043_9 B.subtilis genomic DNA fragment (88 kb). NID: g1945641.

5 atgacaagcaacgaaaaagaaatgggtaaaagtgaattggttagttgttacaggtatgtct
ggagcgggttaaatacattggtgattcaaaagtctcgaagatatgggatttttctgtgtagat
aattttaccacctgtactattacctaattttgtagaattgatggctcaaggaaatccttca
ttgcaaaaaagtagcaattgcaatagattttaagaggtaaggaattatttaaattcttagtt
aaagaaattgatattattaaaaagtcgtaatacgctgatttttagatgttatgttttagaa
gctaaaaactgaaaaaattatttcacgttataaagaatcaagaagagcgacccactaaat
10 gaacaaggacaaagatcattaatagatgcaataaatgaggaacgtgaacatctatcagaa
atccgaagtatcgctaattacgtgattgatacaacaaaattaaaacctaaagaattaaag
caacgcattttcaaagttttatttagatgaaaaactttgaaacatttacaatcaacgtgaca
agtttcggtttcaagcatggtatacaaatggatgctgatttagttttgatgtcagattt
ctacctaataccctactatgtagaggaattgctgctcatttactggttttagatgagccagt
15 tacaattacgttatgaagtggaaagaaacccaaatattttttgataaattaacagattta
ttaaatttatgattcctggctacaaaaaagaaggtaaatcgcaattggttattgctata
ggttgtagcgggtggacaacatcgatcagtcgcattagctaaacgttttagctgaatatctt
aacgagattttgaaataaatgtttatgtgcatcatagagatgcgcattattgaaagtggag
gagagataa

20 Sequence 1144
MTSNEKEMKSELLVVTGMSGAGKSLVIQSLDMGFFCVDNLPPVLLPKFVELMAQGNPS
LQKVAIAIDLGRKELFKSLVKEIDIKSRNDVILDMVFLEAKTEKIIISRYKESRRRAHPLN
EQGQRLIDAINEREHLSEIRSIANYVIDTTKLKPKELKQRIKSKFYLDENFETFTINVT
25 SFGFKHGIQMDADLVDFVRLPNPYVEELRPFTGLDEPVYNYVMKWKETQIFFDKLTDL
LKFMIPGYKKEGKSQLVIAIGCTGGQHRVSVALAKRLAEYLNEIFEYNVYVHHRDAHIESG
ER*

Sequence 1145
30 Contig_0563_pos_0_861,
putative peptide of unknown function
atgaaaaatgaactaacacgcatagaaagttagcgaatcgaatgctaaagcagagctcagt
gcattaattcgcatgaatggcgcacttagtctatcaaatacaacagtttgtaattaatgta
cagacagaaaaatgcgacaacagctcgctgaatttactctcttatcaaacgtatatattaat
35 gttgaagttgaaatttttagttagaaaaaagatgaaattgaaaaaaaacaatatttatata
tgtcgaaacaaagatgttagcgaaagaaatactaaatgatttaggaattttaaaaaaggga
gtttttactcacgatattgatccggatatgattaaagatgatgaaatgaaaagaagttat
ttaagaggggctttcttagcaggtggttctgtaaataatcctgaaacatcttcataatcat
cttgaaattttttcacaaatgaagatcattccgaaggtcttactaaattgatgaatagt
40 tatgaaactcaatgcgaaacatttgaacgtaaaaaaggagattgctgatcttaagaa
gctgaaaaaattttccgactttcttagttgataggtggctatcaagcattgttaaagttt
gaagatgtaagaattgtccgtgatatgcgtaattcggttaatcgtcttgtaattgtgaa
acagcaaatcttaataaaaactgttagcgacgaatgaaacaggttgaaagtatacaatta
attgatgaagaaattgggcttgaaaatttacctgatcggttaagagaagtagcgaagctc
45 agagtagaacatcaagaaatcgtttaaagaattgggtgagatggtttctacagggcct
atatctaaatcaggtACCATT

Sequence 1146
50 MKNELTRIEVDESSNAKAELSALIRMGALSLSNQFVINVTENATTARRIYSLIKRIFN
VEVEILVRKKMKLKKNNIYICRTKMLAKEILNDLGILKKGVFTHDIDPDMIKDDEMRSY
LRGAFLAGGSVNNPETSSYHLEIFSQYEDHSEGLTKLMNSYELNAKHLERKKGSIAYLKE
AEKISDFLSLIGGYQALLKFEDVRIVDRMRNSVNRLVNCETANLNKTVSAAKQVESIQ
LIDEEIGLENLPDLRLREVAKLRLVEHQEISLKELGEMVSTGPISKSGTI

55 Sequence 1147
Contig_0564_pos_1869_2756,
is similar to (with p-value 0.0e+00)
>sp:sp|P37527|YAAD_BACSU 31.6 KD GUANYLYLATED PROTEIN IN DAC
A-SERS INTERGENIC REGION. >gp:gp|D26185|BAC180K_75 B. subtil

is DNA, 180 kilobase region of replication origin. NID: g467326. >gp:gp|Z99104|BSUB0001_11 Bacillus subtilis complete genome (section 1 of 21): from 1 to 213080. NID: g2632267.

5 atgtctaaaatagtaggatcagatcgagttaaaagaggaatggctgaaatgcaaaaaggc
gggtgtcattatggacgtcgtaaatgcagaacaagctaaaattgctgaagaagccggagct
gttgccgtaaatggcattagagcgtgtaccatcagatattcgctgctgctggcggtgttgca
cgtatggcgaatcctaaaatagttgaagaagttatgaatgccgtatcaattccggttatg
gctaaagccagaattgggtcatattacagaagctagagtttagaatcgatgggtgttgac
10 tatatagatgagtcgaagtattacgcctgcagacgaagaatatcatttaagaaaagat
caatttacagttccttttggctgtgctgtaacttaggtgaagcagcagcagcatt
gggtgaagggtcgccgatgttgcgtacgaagggtgaacctggtaggtgaatattgttgaa
gctgtccgctcatatgagacgtgttaattctgaagttagccgcttaacagttatgaatgat
gatgaaattatgacatttgcaaaagatttgggtgcaccttatgaagtattaaacaaatt
aaagataatggacgtcttcctgtagtttaatttgcagctgggtggtgtgctacgcctcag
15 gatgcagcactaatgatggaattaggtgcagatgggtgatttgttggttcaggtatattt
aaatctgaagatcctgaaaaatttgcataagctatcgttcaagctacaacacattatcaa
gattatgagtttaactcgaaaaattggctagtgagctaggtacggctatgaaaggtctagat
attaatcaaatctcactagaagaagaatgcaagagcgtgggttggttaa

20 Sequence 1148

MSKIVGSDRVKRGMAEMQKGGVIMDVVNAEQAKIAEEAGAVVMALERVPSDIRAAGGVA
RMANPKIVEEVMNAVSIIPMAKARIGHITEARVLESMGVYIDSEVLTPEDEEYHLRKD
QFTVPFVCGCRNLGEAARRIGEAAMLRTKGEPTGNIVEAVRHMRVNSEVSRLTMND
DEIMTFKDLGAPYEVLLKQIKDNGRLPVVNFAAGGVATPQDAALMELGADGVFVSGSIF
25 KSEDPEKFAKAIVQATTHYQDYELIGKLASELGTAMKGLDINQISLEERMQERGW*

Sequence 1149

Contig_0564_pos_2969_3316,
is similar to (with p-value 1.0e-33)

30 >sp:sp|P37528|YAAE_BACSU HYPOTHETICAL 21.4 KD PROTEIN IN DAC
A-SERS INTERGENIC REGION. >gp:gp|D26185|BAC180K_76 B. subtilis
is DNA, 180 kilobase region of replication origin. NID: g467326.
>gp:gp|Z99104|BSUB0001_12 Bacillus subtilis complete genome (section 1 of 21): from 1 to 213080. NID: g2632267.

35 atgtttggaacatgtgctggttaattgttcttgcaaaaaatgttgaaaatgagtcctggt
tatttaataaattagatataactgttgagcgaattcattccggtagacaagtcgatagc
tttgaatctgaacttgatattaaagggatagcaaatgatattgagggagttattattaga
gcacctcatatttgctaaagtggataacggagtggaataacttagtaaaagttggaggtaaa
atagtagccgtcaacaaggacaataacctcggtgttctttccatccagaactaactgat
40 gattatcgatatcactaagtattttattgaacacatgattaaacattaa

Sequence 1150

MFGTCAGLIVLAKNVENESGYLNKLDITVERNSFGRQVDSFESELDIKGIANDIEGVFIR
APHIAKVDNGVEILSKVGGKIVAVKQGQYLGVSFHPELTDDYRITKYFIEHMIKH*

45

Sequence 1151

Contig_0564_pos_3912_4925,
putative peptide of unknown function

atggaacgattttgtgtgttaaatcaaatataattcaaatgaatccggttagaagcc
50 aaatttaaaacgagcgtctaaagatcatggaactgatcaggcagatgctcataagcct
gcttgttttaggaccgagcgttaaaacacagacagcttacctatacatgagtttaatttc
tttgaattaaagaaacgcgtccggttttcatctagaaatcgagaatgaacaaaatcgactt
aaatttcagatccttgattactccatcaaacattccctgggttagaaagattgtttagt
agtcgatattcaatcattgcactcaacatcgagaaatctttactcatccagacatgggtt
55 cttgatatcgacaaggaggtactgattacacatatattcaattctacagataagggaatg
tcaatggataaagctacaaaatatgcacttcaattaagggtgattgctcaagaaagctat
cctagtgtcgatagacattcctttctagtcgaaaaattacgcttacttattcaacaatta
aaacaatctattcatctcaacaattagatgatgccatgattcaattagcacaacaa
ctcgattattttgaaaatattcattcgatacctgggtatttggttaagctaagcacagctatg

attattggggagattggtgatattaagcgatttaaatacaataaacaactcaatgctttt
 gttggcattgatatacaacgatatacaatcaggtcatacacactgtagagataccatcaac
 aagcgtggttaataaaaaagcgagaaaacttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgtcgattattactacaaactaagaaagcagcct
 5 aatgagaaacctcataagactgccatcattgcttgataaatcgattatataaaacaatt
 cattatcttgtaatgaatcataaattgtacgattatcaaatgtcaccacattag

Sequence 1152
 MERFCCVNQINYIQMNPLEAKFKTSALRSWKTDQADAHKLACLGPTLKQTDLSLPIHELIF
 10 FELRERVRFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSI IALNIAEIFTHPDMV
 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPSVDRHSFLVEKLRLLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFKSNKQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLFWVIMNIIRGQHHDYDNDHVVDDYKLRKQP
 NEKPHKTAIIACINRLKTIHYLVMNHKLYDYQMSPH*

Sequence 1153
 Contig_0565_pos_776_1195,
 is similar to (with p-value 6.0e-53)
 >sp:sp|P32727|NUSA_BACSU N UTILIZATION SUBSTANCE PROTEIN A H
 20 OMOLOG (NUSA PROTEIN). >pir:pir|C36905|C36905 nusA homolog -
 Bacillus subtilis >gp:gp|Z18631|BSORF17A_2 B.subtilis infB
 -nusA operon. NID: g49314.
 gtgctgtcagaagctgaaagaagtcctaataagagaaatataattcctaatagaacgtatcaag
 gtgtacgtaaataaagttgaacagactacaaaagggtccacaaatttacgtatcaagaagt
 25 catcctggattactaaaacgcttattcgaacaagaagttccagaaatttatgatggact
 gttattgttaaatcagtagcgcgtgaagctggagatcgttctaaaattagcgtgtattct
 gataatcctgatatagatgctgttgccgcgtgtgtaggttctaaaggagcacgagtagaa
 gcggttggtgaagaacttggtggcgaaaaaatcgatatcgccaatgggatgaagatccg
 aaagtatttgttcgtaatgctttaagtcacacaaagttttagaagtaattgttaataaa

Sequence 1154
 VLSEAERSPNKEYIPNERIKVYVKNVEQTTKGPQIYVSRSHPGLLKRLFEQEVPEIYDGT
 VIVKSVAREAGDRSKISVYSDNPDI DAVGACVGSKGARVEAVVEELGGEKIDIVQWDEDP
 35 KVFVRNALSPSQVLEVIVK*

Sequence 1155
 Contig_0565_pos_2337_3152,
 is similar to (with p-value 1.0e-35)
 40 >gp:gp|M24523|BACRTP_3 B.subtilis rtp gene, complete cds and
 proC gene (put.), 5' end. NID: g143477.
 atgaaacttgatattttatgggtgctggaatatggcgaggcaatttttactggaattatt
 aattccaacaatttaaatgcaaatgatatttttaactaataaatccaatgaacaagca
 ttaaaaagcctttgcagaaaaattaggggttaattatagttatgatgatgaagcattactc
 45 aaagatgccgattatgtatttttaggtacaaagcccatgattttgaaaatttagctaatt
 cgtattagagaacacattactaatgataataggtttatttctataatggcagggttatct
 attgattatattcgtcagcagcttaataccaataatccattagctcgtattatgccaaat
 acaaatgctcaagttggacattcgggtactggaataagtttttcaataattttgatcct
 aatctaaaaatgaagtggatgaattaatcaatgcatttgatcagttatagaagtcctcc
 50 gaagaacatctacatcaagttactgcaattacaggaagtgggcctgcatttttatatcat
 gtatttgaacaatatgtaaaagcaggtacagaattaggtttagaacgaaatcaagtcgaa
 gaatctacatgcaatttaattattggaacaagtaaaatgattgagcgttcagacttaagt
 atgtctcaattaaaggaataattacatctaaagggtggtactacacaagctggacttgat
 gcactatctcaatatgatattgtatcgatgtttgaagattgttttaggtgcagctgtgaat
 55 agaagtatggaattatcacataaagaagatgaataa

Sequence 1156
 MKLVFYGAGNMAQAI FTGI INSNLNDIYLTNKSNEQALKSFAEKLGVNYSYDDEALL
 KDADYVFLGTPHDFENLANRIREHITNDNRFISIMAGLSIDYIRQQLNTNNPLARIMP

TNAQVGHSVTGISFSNNFDPKSKNEVDELINAFGSVIEVSEEHLHQVTAITGSGPAFLYH
VFEQYVKAGTELGLERNQVEESIRNLIIGTSKMIERSDLSMSQLRKNITSKGGTTQAGLD
ALSQYDIVSMFEDCLGAAVNRSMELSHKEDE*

- 5 Sequence 1157
Contig_0565_pos_4011_3202,
is similar to (with p-value 6.0e-57)
>sp:sp|P54548|YQJK_BACSU HYPOTHETICAL 34.0 KD PROTEIN IN GLN
Q-ANSR INTERGENIC REGION. >gp:gp|D84432|BACJH642_266 Bacillu
10 s subtilis DNA, 283 Kb region containing skin element. NID:
g2627063. >gp:gp|Z99116|BSUB0013_96 Bacillus subtilis comple
te genome (section 13 of 21): from 2395261 to 2613730. NID:
g2634723.
gtgggcgaggggaacacacatcaaatttttaagacactctattaagtttagggaagatagat
15 catatttttataacacacatgcatggtgatcatatttttggtctccctggacttttaaca
agtcggttcggtttcaagggtggagaaaataaacccttactattatagggcctaaaggatt
caaaattacatagaaacatctttacaactttctgaatcgcatttgaattatccgattacc
tatatcgaaatcaatcaacaattagcgtatcaccacaatggttttactgtacaagctgaa
atgcttaaccatggcatacccttcattcggatcgtattgaagcccaatcacgcctggt
20 acaatcaatgtagaggccttgagaggtattggactagagcctggtccaaaatatcaggaa
gtcaaattacaagaaacgttcgaatataaaggattaatttacaattcggatgattttaaa
ggtaaagctaaacctggtccaattatcagtatatttggtgatacaaaaccgtgtgaaaat
gaatatgaattagcaaaagaactcagatttaatgattcatgaagcaacttacattgaagga
gataagaaacttgctaataattaccatcatagtcatatagacgatgtatttaattctaat
25 aagcaagctaattgtaataaaaagtcttatcactcatatcagtaacagatataacattgat
gaagttacatcaatatacaatgagttatcccttgatcaaaacttctccacatttttatttc
gttaaagattttgatactttcaaaatataa

- Sequence 1158
30 VEGTQHQLRHSIKLGKIDHIFITHMHGDHIFGLPGLLTSSRSFQGGENKPLTIIGPKGI
QNYIETSLQLSESHLNPITYIEINQQLAYHHNGFTVQAEMLNHGIPSFQYRIEAPITPG
TINVEALRGIGLEPGPKYQEVKLQETFEYKGLIYNSSDFKGAKEPGPIISIFGDTKPCEN
EYELAKNSDLMIHEATYIEGDKKLANNYHSHIDDVFNLIKQANVNKSLITHISNRYNID
EVTSIYNELSLDQTSFPHYFVKDFDTFKI*

- 35 Sequence 1159
Contig_0565_pos_2176_1418,
putative peptide of unknown function
gtgataggcaaacactttattataactggagcaacgagtggttaggttttgcaataacc
40 aatgaattacttcaaaagaggggcccattgtactatacttgcaagaaatatagataagttc
aatcgaatcaaagaaaactattttaaacctgaacatatcaatgtgattaaatgtgattta
atgcaacgaaaagatatattgaatcattacaaaaatttttaatacacctataaatggtttc
atctacagttcaggtgttgatattttaagtctataagtgagcattcaactcgtgaagta
gtagaaacttacgaggttaattcttacaatttttaatttggtatacaaaagtgttcaacca
45 caattagtaaaagcagcatatctcgttggtatatctagtcaagctgctcttgtttcacag
gctaattgcggcacattacggtgcatcgaaagcagggttagcgccgttcttaatgcattg
agattagaacaaccggaattaaaagtgtcaatgtacagcccggtccaatagatacacca
ttccaaaaaaacgcagatcctactctaaagtattttaaaaattatagacacatgatgata
caacctcaacaacttgccaagcaaatagtggaaggaataataactaaataaaattgaaatt
50 aatcaaccatcatggatgcaataatgcttaattttatcaattatgtccacgtacacta
gaaaaattatgtccaaatctattttaaaaataaagtttaa

- Sequence 1160
55 VIGKHFIITGATSGLGFAITNELLQRGAHVITILARNIDKFNRKENYFKPEHINVIKCDL
MQRKDIESLQKFLNTPINGFIYSSGVGYFKSISEHSTREVVETYEVLNLTNFNLLYKVIQP
QLVKAAYIVGISSQAALVSQANAHHYGASKAGFSAVLNALRLEQPELKVLNVQPGPIDTP
FQKNADPTLKYFKNYRHMMIQPQLAKQIVEGIILNKIEINQPSWMQIMLFYQLCPRTL
EKLCPNLFKNKV*

Sequence 1161

Contig_0566_pos_1171_1491,

putative peptide of unknown function

5 atgaatgtacagtttaagaaagggtgctttagaattaattgttctgctaattattaaaaaa
gaagatcagtatggttattcacttgtacaaaatatctccagatatatgaccatagctgaa
ggtagcagtttatcctctgctaaggcggttggtaaaagtggggaactgagtacgtattat
caaccttcaactgaagggtccgtctcgaaagtattatcaattaactcaacagggggctgag
agagttaatcaattagaggaggattggaaattggttacggaagctgtagaacatttcatt
10 gaggagagtgagaatgaatga

Sequence 1162

MNVQFKKGALELIVLLIIKKEDQYGYSLVQNISRYMTIAEGTVYPLLRRLVKSGELSTYY

QPSTEGPSRKYYQLTQQGAARVNQLEEDWKLFTAEVHFIEESENE*

15 Sequence 1163

Contig_0566_pos_1536_2048,

putative peptide of unknown function

atgaataaagaagaaaaagaagatattttgaatgaatacgcacatcttttatagcgga
cagcaagagggaaagtctgaatcagacgtgtgtaagaattaggtaatccaaaattaata
20 ggtaaggaacttacagctacttccagtgtagaaaatgcacatcaaaaagtgtcgtaatg
aatatttcatccgcaattgttagcagtaatggggttaagttgcttaactttttattggt
ataataaccagcttttttatgcattttgctcgatttaaccttcatcttttactctagct
tcactagctgcaccattgatgttgctcattaaaggaattatggatgggtttcattccatt
atcttatatgacgcatttatgactgggttaatggttggttgactcgacttgcagtg
25 gtgacttactatctcattaaagtggctatttgatgtgactatgaaatatctaaaatggaat
atctctattgtcaaaggaagtgtacaatcatga

Sequence 1164

MNKEEKEDILNEYDTHFYSGQQEGKSESDVCKELGNPKLIGKELTATSSVENAHQKVSLM

30 NISSAIVAVMGLSLNFFIVIIPAFLCILLVLTFFIIFTLASLAAPLMLLIKIMDGFHSI
ILYDAFMTGLMFGVLVLA VVTYYLIKWLFDVMTMKYLKWNISIVKGSVQS*

Sequence 1165

Contig_0566_pos_2105_2791,

35 putative peptide of unknown function

gtgggagtttatgcacaaaataaaaaattgagtaagacaatcaatataataatcaaaaa
acaaatttaatgaaaaactatgatgataatactgtgaaaagtattttacattgatgaaaa
gtaagtgatataactgtgaaaaaaggtaaacatttttcggttaagtccaaagggaatgac
aaaaatttaaacgttaactagcaaggtgaacaatacaacgttgggttaattacagagcgtaa
40 acaagtccacatattaattttagaatacaaggtaaaagttagtaatcacattacgattaca
gtacctaaatatattaaaaacatagatatataaaactaatgccggggatttaaatattggt
ggagtaaatagtgccacaggaagatttgatgctgaatctggagacattaaagttcaaaaa
ggacgatataaaaaagggtgacacttcataatgaggatggggatattcaaatgaaacaatta
gacctgatattcctttacgtattaaaaatgaagaaggggatataaaactgaattataaa
45 aaagaacttcatacacacccaaatcatcactcgtaatgaagaaggggaaacagacatcgat
catcggtgtgttatataatagtaaaagttgaaaatgaaaataataaagtgaattaatcaat
gaaaatggagatatataaagtaaaataa

Sequence 1166

50 VGVYAQNKKLSKDNQYNNQTTNLMKNYDDNTVKSIIYIDGKVSDITVKKGKHFSVSKGND
KNLNVTSKVNQRWVITERQTSFPHINFRIQGVSNHITITVPKYIKNIDIKTNAGDLNIV
GVNSGTGRFDAESGDIKVQKGRYKKVTLHNEDGDIQMKQLDPDIPLRKNEEGDINLNYK
KELHHTQIITRNEEGETDIDHRVLYNSKVENGNKVKLINENGDIKV*

55 Sequence 1167

Contig_0566_pos_3859_4332,

putative peptide of unknown function

atggtatggcatttaatttttatgattcctacaattattgggttactcttttgaatgttc
tgtttaatatcaagtgaacttttaaaagttaaaggatttctcttgggtgggtaggaatt

tttatcttttctttaatttatattttgatttatagtttagttacattaatacctaattgtt
gcatattatcaagaagattccatgatcgctcaatgacgatgactcttccgattatttt
tatgttttctactgtcattgtatcagggttttaattattaccaaatatagataattcagcg
gtattaatatttatgggaattatctgcttaactctactggattgggtcgatattaatatta
5 gtattgacttgcttagatagtaaaacagagtcctaataaatatggaccaagtccaaagtac
aatcgtaacgagacaaatttccatggtgataatgctaattccagttgataaataa

Sequence 1168

MLWHLIFMIPTIIGYSFGMFCLISSETFKSKGFLLLGVGIFIFSLIYILIYSLVTLIPNV
10 AILSRRFHDRSMTMTLPPIIFYVFTVIVSGFNYPNIDNSAVLIFMGIICLIYWIGSILIL
VLTCLDSKTESNKYGSPKYNRNETNFHGDNANPVDK*

Sequence 1169

Contig_0566_pos_7514_8113,
15 putative peptide of unknown function
atgatggtaaaggttattcacacttatgatgccaatcatcgatggtccgtacaatatgag
gcaaaatcaacaaaaaacagttatccttcaaatcatgtttattttaatttgaat
cgtgataacaatgttgctataaccactgtataaatagttcagcattaaaaatgtatatg
ttaaataataaacatattgttaagaggggcaatctcttgatttacatcgattattggat
20 acaataaaagtctattttaaagatatttttgaaagtgaacaatgaaactttgcaacaaca
attaatcattataatggcattgatcctttcgaatttggtggaaatgagcttaccatt
gataattctgaatttgaacttaataaagactgatatgcctcattttgtaattgttacg
tttaatgatccacaagtttggaacaatgattttaataattataaaggcgcatcttgattt
tccatcgaaactcaatatatgcctaataatgatataaatatgtacggcgccaaagctcagtt
25 attttagaggcagatacattatttacatctaaaaacaagttttcaaattcatgaaaagtag

Sequence 1170

MMVKVIHTYDANHRWSVQYEAKSTKKTVFNPNSNHVYFNLNRDNNVVYNHCINSSALKMYM
30 LNNKHIVKEGQSLDLHRLDNTNKVYLKDI FESDNETLQQQINHYNGIDHPFEFGGNETI
DNSEFELNIKTDMPHFVMTFNDPQVWNNDFN IYKAHSGFSIETQYMPNDINMYGAKAQS
ILEADTLFTSKTSFQIHEK*

Sequence 1171

Contig_0566_pos_6470_4905,
35 is similar to (with p-value 1.0e-95)
>gp:gp|X89408|BSARAABD_2 B.subtilis DNA for araA, araB and a
raD genes. NID: gl924929.
gtgatattagctgacacatccaacggacatatcatatcaagatatgaggaagactatgcg
40 aacggaacttatatgaactcattatatgataaaccgttacctgaaaactacttcttaca
aatgctgacgactatttacaattcttgaacaaggcgttcaatttgattagaagatagt
aaagttaataaaaacgatgtggttggaattggagtcgactttacaagcagtaacaattatc
tttctcgatgaacaatttgaaccgcttcacgtcatgaagatttaaagacaaatccacac
gcgtacgtaaaattatggaacatcatggagctcaagatgaggcaaaactatatgattcag
45 atgagtaagaataaaaaattggttagattattatggctcaagcgtaaatagcgaatggatg
ataccgaaaatcctggaagttaaacaatgaagcaccagaaatacttagaagagcacggtat
ataatggaagctggagattacatcactagtataactaacaattcaaataacgatcaaat
tgtggtattggttttaaaggtttttgggacaatgaagctggatttaattacgacttcttc
catagcgtggatcctgatttacctaaaatcgtcaaagaaaaatgtgaagcgccaatcata
50 tcaattgggaaagtgcaggctcgtttatgtaaagactatcaacaatatgggggctttct
caatatgtccagggtttcaccttttatcatagatgcacattctggcgtcttaggtgttggg
gcaatagaagctggagaattcactgcagtcattggtacaagtaacttgcatctcatgcta
gattcaagacaagtaaccatttcttcaataactggctcagttaaaaatgctattatacct
ggattatatgcctatgaagctgggtcaaccagctgtcggtgatttggttgaaactcaaag
55 aaccaagcacctaaacatattgtagatcaagcaaatgaacatcatatgcatgtgcttaac
tatttagaggaattagcaagtcacattagaatagaagaacaacatgttggttggttagat
tggttgaaatggaaatcgtagtatacttagtaaatagtcactaactggaagtatcttgggt
cttacacttcaaacaccgtatgaaatgattcatcgagcatatattgaagctacagcattt
ggaacaaaattaattatgaacaatttgaagataatcatattcctgttcatacagtgat

gcgtctggtggtatcccacaaaagagtaaattactcgttgaaatttatgcaaattgtttta
 aataaaagggttgctgcgtcatagattcatctaattgcttcagcattaggtgcagcgatgtta
 ggtgcaaattgttggaatgcatatagtagacattaaaagaggcggcattatctatgaagcaa
 cctatagcttatatacaagaacctgaaatccaaaaagttcaagcttataaaccactctac
 5 cataaatattgtgaactacatgattttatttagatcgtcaatatcctgaattatcatatttg
 atttaa

Sequence 1172

VILADTSNGHIISRYEEDYANGTYMNSLYDKPLPENYFLQNADDYLQILEQQGVQFVLEDS
 10 KVNKNVVGIGVDFTSSTIIFLDEQFEPLHRHEDLKTNPAYVKLWKHHGAQDEANYMIQ
 MSKNKNWLDYYGSSVNSEWMIPKILEVKHEAPEILRRARYIMEAGDYITSILTNSNIRSN
 CGIGFKGFWNEAGFNDFHFSVDPDLPKIVKEKCEAPIISIGESAGRLCKDYQQIWGLS
 QYVQVSPFIIDAHSGVLGVGAIEAGEFTAVIGTSTCHMLDSRQVPISSITGSVKNAIIP
 GLYAYEAGQPAVGDLFEYSKNQAPKHIVDQANEHHMHVNLNLEELASHIRIEEQHVVDL
 15 WLNNGRSILSNSHLTGSIFGLTLQTPYEMIHRAYIEATAFGTKLIMKFEDNHPVHTVY
 ASGGIPQKSKLLVEIYANVLNKRVVVIDSSNASALGAAMLGANVGNAVSTLKEAALSMKQ
 PIAYIQEPEIQKVQAYKPLYHKYCELDLDRQPELSYLI*

Sequence 1173

20 Contig_0566_pos_854_495,
 putative peptide of unknown function
 atgatagaatttgatgcaattaccacattatgtttggcatgtgttttatatttaattggt
 caaacaataatcaaccatgtttctatttttaaggcgaatctgtattccagcacctgtcatt
 ggtggtcctttatttgcaatatttagtggtatatttagattcatttaatatcgttaaaata
 25 aaacttgattcggcggttcattcagaatttctttatgctcgtttctttactacaattgga
 cttggcgcatcattaaaactattcaaaattggcggaaaagtcattgttgatagcgagaatg
 tctccatttttaggattttggacaaccattaacgcattgtccatccttagcaccttga

Sequence 1174

30 MIEFDAITTLCLACVLYLIGQTIINHVSILRRICIPAPVIGLLFAILVAILDSFNIVKI
 KLDSAFIQNFMLAFFTTIGLGASLKLKFKIGGKVMLIARMSPFLGFWTTINALSISLAP*

Sequence 1175

35 Contig_0568_pos_6584_7120,
 is similar to (with p-value 3.0e-26)
 >gp:gp|AF024506|AF024506_1 Bacillus subtilis SecDF protein (secDF) gene, complete cds. NID: g3220155.
 40 atgataatcaaaccaataattacaattaaaactaagtgaataaagtggttagctaatt
 ttaacaagaagtttaaccttcatatgatgtttttaaatcatgtacatctttaccttcatta
 atatcatgtctatccttcttcttaacaccaaataaccagattgttttttaagaagttt
 gaagataccagtaagtataacaacctcttgataagaatactgcggttacaatatcatt
 aaaatacctaagagtaacatggttgcgagcctttgactgaactttctccaaagaagaaa
 45 agcacagctgcagcgatgacagttgttaagttggaatcaaatatagttagaatgaactt
 ttatttgcttttgaatacgttggttaagcgtgcgtccaattcttagttcatctttaata
 cgttcatacattatgatattggcatcgacagccatacctacacctaataatgcccgc
 aatccaggtagagtttaatacacctgatatgaaattgaatgcgactaaagttaaataa

Sequence 1176

50 MIIKPIITIKILSEISGLANLTKFNLSYDVKFSCSLPSLISCLSFFLTPNNQYCFLLKKF
 EDTSDNNPLDKNTAVTNIKIPKSNMVAKPLTELSPPKKKSTAAAMTVVKLESNIVKNE
 LFAFEYACLSVRPILSSSLIRSYIMILASTAIPTPKTNAANPGRVNTPDMKLNATKVK*

Sequence 1177

55 Contig_0568_pos_7956_5731,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF024506|AF024506_1 Bacillus subtilis SecDF protein (secDF) gene, complete cds. NID: g3220155.

atgacgtataagaatgtagttaaaaatgttaatttaggtctagatttgcaaggtgggttt
 gaagtcctcttccaagtagatcctttaataaaggagataaaattgataaaaaagcactt
 caagctacatctcaaacattagaaaatcgtgtaaatgttctaggtgtatcagaaccgaaa
 atacaaatcgaagatccaaatcgaattcgtgtacaattagcaggtatcaaggatcaagca
 5 caagcgcgtaaattattatcgacacaaagctaatttaacaattagagatgctgaagatcat
 gttttaatgtctggttcagacattaaacaaggctctgctaacaagaatttaaaacaagaa
 actaatcaaccaacagttacatttaaaagtaaaaagtaaaagataaatttaagaaagtaact
 gaaaagatttctaaaaaacgtgacaatgtcatggtagtttggttagatttcgaaaaaggc
 gatagttacaagaaagaagctaaaaagcaacaagaaggtaaaaagcctaatttatatct
 10 gcagcagagtgtagaccaacctattaattctagtagtgttgaaatttcaggtggcttcaat
 gggaaaaaagggtgtgaagaagcgaacaaatagctgagttattaaatgccggctcatta
 ccagttgattttaaaagaaatttactctaactctggttggtgcacaatttggtcaagatgct
 cttgataagaccatgtttgcatcaattgtaggtatagcatttaatttatttattttagctt
 ggtttctatcgtttgcctgggttagttgcaatcattgccttaaccacttatatttattta
 15 actttagtcgcattcaatttcatatcaggtgtattaactctacctggattggcggcatta
 gttttaggtgtaggtatggctgtcgatgccaatatcataatgtatgaacgtattaaagat
 gaactaagaattggacgcacgcttaaaacaagcgtattcaaaagcaaaataaaagtccattc
 ttaactatatttgattccaacttaacaactgtcatcgctgcagctgtgcttttcttcttt
 ggagaaagttcagtcgaaaggcttcgcaaccatgttactcttaggtattttaatgatattt
 20 gtaaccgcagtatcttatcaagaggggtgttatcattactggtatcttcaaacttcttt
 aaaaaacaatactggttatttgggtgttaagaagaaggatagacatgatattaatgaagg
 aagatgtacatgatttaaaacatcatatgaaagggttaactttgttaaattagctaag
 ccacttatttcacttagtattttaattgtaattattggtttgattatcatttcaatattt
 aaattaaacttaggtattgatttctcatccggaacaagagcagatattcaatctaaaaat
 25 gctataacacaagcacaggttgagaaaactgtaaaatcagttggattggaaccagatcaa
 atacagattaatggtagtggaataaaaaatgccacagttcagtttaaaaaagatttatca
 cgtgaggaagacaataaattaagtgttaagggtgaaatctgaatttgagataatccacaa
 attaataccgtttcacctctcataggccaagagctagctaaaaatgctgtaactgcatta
 atacttgcttctataggcattattatctatgtttcactaagatttgaaatggcgtatgggt
 30 ctatctctgtacttgcatattatcatgacgtatttatcatcattgcaatcttttagtttg
 tttagattagaagtagatttaacatttattgcagcagatttaactatcgttggttattca
 atcaatgatacaaatcgtaactttcgaccgtgttcgagaaaatctgcataaagttaaagta
 attacgcatactgatcaaattgatgatatagtcaaccgctctattagacaaactatgaca
 cgttctattaatacagtggtgactgtagttgtagttgtagttgcaatattaatattaggt
 35 gcaccaacaataatttaatttctcttttagcattactaattggattattatctggtgtattc
 tcgtcaattttcattgctgtaccattatggggcatgcttaagaaacgacagtttaaaaag
 acaaaaaataataaatttagtagtacacaaagagaagaatctaacgatgaaaaaatctta
 gtttaa

40 Sequence 1178
 MTYKNVVKVNVNGLDLQGGFEVLFQVDPLNKGDKIDKKALQATSQTLENRVNVLGVSEPK
 IQIEDPNRIRVQLAGIKDQAQARKLLSTQANLTIRDAEDHVLMSGSDIKQGSARKQEFKQE
 TNQPTVTFFVKVSKDKFKKVTEKISKKRDNVVVWLDFEKGD SYKKEAKKQEGKKPKFIS
 AASVDQPINSSSVEISGGFNGKKGVVEAKQIAELLNAGSLPVDLKEIYSNSVGAQFGQDA
 45 LDKTMFASIVGIALIYLFMLGFYRLPGLVAIIALTTYIYLTIVAFNFISGVLTLPLAAL
 VLGVMAMDANIIMYERIKDELRIKQAYSKANKSSFLTIFDSNLTTVIAAAVLEFF
 GESSVKGFATMLLLGILMIFVTAVFLSRGLLSLLVSSNFFKKQYWLFGVKKKDRHDINEG
 KDVHDLKTSYERLNFVKLAKPLISLSILIVIIGLIIISIFKLNLDIFSSGTRADIQSKN
 AITQAQVEKTVKSVGLEPDQIQINGSGNKNATVQFKKDLREEDNKLSAKVKSEFGDNPQ
 50 INTVSPLIGQELAKNAV TALILASIGIIYVSLRFEWRLSSVLALLHDVFIIIAIFSL
 FRLEVDLTFIAAVLTIVGYSINDTIVTFDRVRENLHKVKVITHTDQIDDIVNRSIRQMT
 RSINTVLTVVVVVVAILILGAPTIFNFSLALLIGLLSGVFSSIFIAVPLWGM LKKRQFKK
 TKNNKLVVHKEKKSNDKILV*

55 Sequence 1179
 Contig_0568_pos_2695_2177,
 is similar to (with p-value 2.0e-90)
 >gp:gp|D76414|D76414_1 Staphylococcus aureus gene for histid
 yl-tRNA synthetase, ppGpp hydrolase, lytic enzyme, complete

cds. NID: g2580431.

atggatttaaaacaatatgtttcagaagtaaaagattggccttcagcaggtgtaagcttt
aaggatataacaactataatggataatggtgaagcttatggatatgctacggatcaaatt
gttgaatacgcacaaaggaaaaaataatagatatagtagttggtcctgaagccagaggattc
5 ataatagggtgtccagttgcttactcaatgggtattggatttgctccagtagctaaagaa
ggaaaactacctcgagaagttattcgttatgaatataatttagaatatggaactaacgta
ttaactatgcataaagacgcgattaaaccaggacaacgagttttaatcactgatgattta
ctagctacaggtggaactattgaagctgcaataaagcttggtgaacaattaggtggtata
gtttaggtattgcttttattattgaacttaaatatttgatggaattgataaaaataaaa
10 gattatgatgtgatgagtttgatttcatatgatgaataa

Sequence 1180

MDLKQYVSEVKDWPSAGVSFKDITTIMDNGEAYGYATDQIVEYAKEKNIDIVVGPEARGF
IIGCPVAYSMGIGFAPVRKEGKLPREVIRYEYNLEYGTNVLTMHKDAIKPGQRLITDDL
15 LATGGTIEAAIKLVEQLGGIVVGIAFIIELKYLNGIDKIKDYDVM SLISYDE*

Sequence 1181

Contig_0568_pos_0_1656,

is similar to (with p-value 0.0e+00)

>gp:gp|D76414|D76414_2 Staphylococcus aureus gene for histid
yl-tRNA synthetase, ppGpp hydrolase, lytic enzyme, complete
cds. NID: g2580431.

gtgaataacgagtatccatatagcgcggtgaggtgctttataaagctaaatcatattta
tcaacaagtgaatatgaatatgttctcaaaagtatcatatagcttatgaggcacataag
25 ggtcaatttagaaaaaatggcttaccttatattatgcacccattcaagttgcagggatt
ttaacagagatgcgttttagacggaccactattgtcgctgggtttctacatgatgtgatt
gaagatacttcttatacatttgaagatgttaagatatgtttaatgaagaaattgcacga
atagtagacggagtaactaaacttaaaaagggttaagtatcgctctaaggaagaacaacaa
gcagaaaatcatcgtaaaactatttattgctattgctaaagatgtacgcgtaatttttagtg
30 aagtttagcagatcggcttcataatatgagaactttaaaggcaatgccaagagagaagcag
gtaagaatctctaaggaaaccttagaaatctatgctcctttagcacatcgctcgggaatc
aacacgataaagtggaacttgaagatacagcgctgagatatattgacagtgtgcaatat
ttccgcatcgtaaatcttatgaagaaaaacgtagtgaacgcgaagcttacattacaat
gcaatcaataaaaattaaaaacgaaatgactaaaaatgaatcttctcgggcgaaattaacggt
35 agaccaaaacatatattacagtataataccgcaaaatgataaaacaaaaaagcaattcgat
caaataatttgatttgcttgcatacgtattatagtttaattcgataaatgattggtatgag
acacttggttttagttcatatatttgaaaccgatgcctggacgttttaagattatata
gctatgcctaagcaaaatattgatcaatcactacataccactgtagttggacccaatggc
gatcctttagaatacaaaattagaacgcacgaaatgcatgaaatcgctgaacatgggtgtt
40 gctgcacattgggcttataaagaaggtaagacagtttaatacaaaaaacacaggattttcaa
aataagcttaatttggttaaaagaacttgctgaaaccgaccatacttctgcagatgcgcaa
gaatttatggaatccttaaaatatgatttacagagcgataaggtatatgcatttactcca
gctagtgtattatagagttaccttatgggtgcagtaccaattgattttgcttatgcaata
cacagtgaagtaggaaataaaatgattgggtgctaagggttaattggtaaaatcgtaacctata
45 gattatgttctacaaactgggtgatattatagagattcgtacaagtaaacattcttacggt
ccaagttagagactgggtgaaatgtgaaatcttctagtgcacaaagtaaaatcaaaagt
ttctttaaaaaacaagatcggtccttctaataattgaaaaaggtaaatatttggtagaagcg
gagattaaagaacaaggattccggtgttgaagatatcttaactgagaaaaatttagaagtc
gttaatgaaaaatatcattttgctaattgatgaagatttgtagcgagctgttggttcggt
50 ggtgttacatcaatacaaatcgtaataaattaAGA

Sequence 1182

VNNEYPSADEVLKAKSYLSTSEYEVVLKSYHIAIEAHKGQFRKNGLPYIMHPIQVAGI
LTEMRLDGPITVAGFLHDVIEDTSYTFEDVKDMFNEEIARIVDGVTKLKKVKYRSKEEQQ
55 AENHRKLFIAIAKDVRVILVKLADRLHNMRTLKAMPREKQVRISKETLEIYAPLAHRLGI
NTIKWELEDALRYIDSVQYFRIVNLMKKKRSEREAYITNAINKIKNEMTKMNLSGEING
RPKHIYSIYRKMIKQKKQFDQIFDLLAIRIIVNSINDCYATLGLVHTLWKPMGRFKDYI
AMPKQNMYSQSLHTTVVGPNGDPLEIQIRTHEMHEIAEHGVAAHWAYKEGKTVNQKTQDFQ
NKLNLWKELAETDHTSADAQEFMESLKYDLQSDKVYAFTPASDVIELPYGAVPIDFAYAI

HSEVGNMIGAKVNGKIVPIDYVLQTDGDIIEIRTSKHSYGPSRDWLKIVKSSSAKSKI
 FFKKQDRSSNIEKGKFMVEAEIKEQGFRVEDILTEKNLEVVNEKYHFANDEDLYAAVGFG
 GVTSIQIVNKLK

5 Sequence 1183

Contig_0569_pos_4130_4513,

is similar to (with p-value 3.0e-45)

>sp:sp|P52026|DPO1_BACST DNA POLYMERASE I (EC 2.7.7.7) (POL
 I). >gp:gp|L42111|BACPOL_1 Bacillus stearothermophilus DNA p
 olymerase I (pol) gene, complete cds. NID: g806280.

10 atgtcagacattgttaaagatgcaaaagcacaagggtatgtggaaacactacttcatcgt
 cgtcgatacattcctgatataacaagtagaaacggttaatttaagaagttttgcagaaaga
 acagcaatgaatacacccatacaaggtagtcagctgacataataaaattagcaatggtt
 aaattcagtgaaaagattaaagaaactaaatatcatgctaagttattattacaagttcat
 15 gatgaactcatatttgaaataccaaaatcagaagtagaagatttttagtaaattttagaa
 gaaattatggaacaagcattagtgtcgtacaccttaaaagtagattcgaattatggt
 gcaacatggtacgatgctaataa

Sequence 1184

20 MSDIVKDAKAQGYVETLLHRRRYIPDITSRNVNLSFAERTAMNTPIQGSAAIIKLAMV
 KFSEKIKETKYHAKLLLQVHDELIFEIPKSEVEDFSKFVEEIMEQALVLDVPLKVDSNYG
 ATWYDAK*

Sequence 1185

25 Contig_0569_pos_5426_6028,

is similar to (with p-value 2.0e-31)

>sp:sp|Q55515|Y553_SYNY3 HYPOTHETICAL 22.5 KD PROTEIN SLR055
 3. >gp:gp|D64006|SYCSLLLH_95 Synechocystis sp. PCC6803 compl
 ete genome, 25/27, 3138604-3270709. NID: g1001291.

30 gtgattgggataactggtggtattgccactggaaaatcaacagtttcagaattattaaca
 gcatatgggttttaaaatcgtagatgctgatattgcttcacgcgaagcagttaaaaaaggc
 tctaagggtcttgaacaagttaaagagatttttggggaagaagcaattgacgaaaatggt
 gagatgaatcgtaaatatgtaggagagatagtttttaatcatcctgacttacgcgaggct
 cttaatgaaatagttcatcctattgtaagagagataatggaacaagagaaaaacaattat
 35 ctagaacatggatatcatgtaattatggatatcccattgtgtgacgaaaatgaactacaa
 gatactgtagatgaagtttgggtggtttatacatctgaaagtattcaaactcgatcgttta
 atggagaggaataatttatcattagaagatgctaaagcacgtgtttatagtcaaatatct
 atagataaaaaaagtaggatggcagatcatgtgatagataatctaggtgataaattagaa
 cttaaacagaatttacaaaaattacttgaagaagaagggtatattcaatcgagagagtga
 40 tag

Sequence 1186

VIGITGGIATGKSTVSELLTAYGFKIVDADIASREAVKKGSKGLEQVKEIFGEEAIDENG
 EMNRQYVGEIVFNHPDLREALNEIVHPIVREIMEQEKNNYLEHGYHVIMDIPLLYENELQ
 45 DTVDEVVVVYTSESIQRILMERNNLSLEDAKARVYSQISIDKSRMADHVIDNLGDKLE
 LKQNLQKLLEEGYIQSESE*

Sequence 1187

Contig_0569_pos_6872_7897,

is similar to (with p-value 2.0e-96)

>pir:pir|JS0164|JS0164 glyceraldehyde-3-phosphate dehydrogen
 ase (EC 1.2.1.12) - Bacillus stearothermophilus >gp:gp|M2449
 3|BACGAPDHA_1 B.stearothermophilus glyceraldehyde-3-phosphat
 e dehydrogenase gene, complete cds. NID: g142951.

55 atggcaacgaatattgcaattaacggtatgggtagaataggtagaatggtgttacgaata
 gcactaaataataaaaaatttaaatgttaaagcgattaacgctagttatccacctgaaaca
 attgcacatttacttaattatgatacgacgcatggagtttatgataaaaaagttgaaccg
 attgaaagtggattaaagtgaatggacatgaaattaaattactttctgatcgcaatcca
 gaaaatttaccatggaatgagatggatattgatgttggtatagaagcgacaggtaaattt

aatcacggagataaagcagttgctcatattaatgcaggtgctaaaaaggtattactcact
 ggaccgtctaaaggtggagacggttcaaagtattgttaaaggagtcaatgataatcaactt
 gatattgatacatagatatttttagtaatgcatcttgactactaattgtatcggacca
 gttgcaaaaagtcctcaatgataaatttggaatcataaatggctgatgacaactgttcat
 5 gcaataacaaatgatcaaaaaatattgataatccacacaaagatttaagaagagcacgt
 tcttgtaatgaaagtattattccaacgtcaacaggtgctgctaaagcacttaagaagta
 ttgcctgaagttgaaggtaaacttcatggaatggctttaagagtaccaacaaaaaatgtc
 tctctcgttgatttagttgttgatttagaacagaatgttacagttacacaagttaatgat
 10 gcatttaaaaaatgccgatttatcaggtgttcttgatgttgagaagctccttagttct
 gtagactttaacacaaatcctcattcagcaattattgattctcaatctacgatggttatg
 ggacaaaataaggtgaaagtattcgcttggtatgataatgaatggggttattcgaataga
 gttgttgaaagtagctgacaaaattggacaattaattgatgataaagcaatggtaaaagcc
 atttaa

15 Sequence 1188
 MATNIAINGMGRIGRMVLRIALNNKNLNVKAINASYPPETIAHLLNYDTHGVYDKKVEP
 IESGIKVNQHEIKLLSDRNPENLPWNEMDIDVIEATGKFNHGDKAVAHINAGAKKVLTT
 GPSKGGDVNMIVKGVNDNQLDIDTYDIFSNASCTTNCIGPVAKVLNDKFGIINGLMTTVH
 AITNDQKNIDNPHKDLRRARSCNESIIPTSTGAALKKEVLPEVEGKLHGMALRVPTKNV
 20 SLVDLVVDLEQNVTVTQVNDAFKNADLSGLDVEEAPLVSVDFTNPHSIIIDSQSTMVM
 GQNKVKVIAWYDNEWGYSNRVVEVADKIGQLIDDKAMVKAI*

Sequence 1189
 Contig_0569_pos_8888_10258,
 25 is similar to (with p-value 1.0e-51)
 >sp:sp|P07908|DNAB_BACSU REPLICATION INITIATION AND MEMBRANE
 ATTACHMENT PROTEIN. >pir:pir|B26580|B26580 replication init
 iation protein - Bacillus subtilis >gp:gp|AF008220|AF008220_
 191 Bacillus subtilis rrnB-dnaB genomic region. NID: g229313
 30 5. >gp:gp|M15183|BACDNAB_2 B.subtilis dnaB gene, encoding th
 e replication initiation and membrane attachment protein, co
 mplete cds, clone pdnaB12. NID: g142862. >gp:gp|Z99118|BSUB0
 015_164 Bacillus subtilis complete genome (section 15 of 21)
 : from 2795131 to 3013540. NID: g2635200. >gp:gp|Z75208|BSZ7
 35 5208_1 B.subtilis genomic sequence 89009bp. NID: g1769994.
 atggggttacaaacctatgaatatggctctaaaaccacaagatggatttgaggtgattaca
 catttcgaattcacctcacacatttagatattttaaatcgactattcacccctttaatc
 ggagttgaattcattggactctatcattttatgagtcgaattcatagataaaagtcaacaa
 ctcgggttaacgcatttatattcatgaatgaactaaaaattaaactatttagatttcagg
 40 gagcaaatggacaatttagaggctattggattgattaaaacatttgtaaggcatgaagaa
 aagtactctcactttgtttatgagtttaattcagcctccaacagcctatcaattttta
 gatcctatgttatcagttttttatttagtgaggttgataaaaaacggttatcaagcactt
 aatcttatttcgaaaaagatgagaaaagatttaagcaaatatcaacagacaactagaaaa
 ttacagaagtattcaacgtacctaataaaggatgaatgttctgatcaaatatttaag
 45 caaatcaaacactatgatggtatagatttatctaatagaacttttgattttgaaatggtg
 agacagatggtgaaccatcattttatttagtaatgaaattatcgataaagaagctaagaat
 ttgattatacaacttgcgacactttatggaattactgaagatggtatgaaaaatgttata
 ttaagttccattaccagtgcacaacaattatcttttgaaagaaatgcgtaagaaagctaga
 acttattacctgattgaacatgataatcaattacaaaaattagagcatcaaacaaataaa
 50 attaacgatgaaaaaaaagatcgacaagcggaagatacaacaaatgattggttacaactg
 ctgatgaaacaagtcgattgatatttagcaagttggtctgattcggaaacctacacag
 tcgcaaaaagatgatagaagaattgattaaccgtgaaaaaatgaattttggtgtaatc
 aatatacttttacagtttggtatgttaaaagaagatatgaagttgccaaaatcttatatt
 tttgaaattgcttccaactggaagaaaattggtatttcaaatgcaaacaaagcatatgaa
 55 tatgcattacaagtttaacacctaataattacgaaacacattctaataataaacgcacag
 aacaatcgtggaagacaaaatcaatttttatccaaagaaaagacacctaataatggcttcaa
 aatagggacgatcaagaagaaaaataaagaaataaatgatgacactctcgaagaagatcga
 caagcatttcttgaaaagtttaatacaaaagtgaagagggaagataactaa

Sequence 1190

MGLQTYEYGLKPQDGFVITHFEFTSQHLDILNRLFTPLIGVESIGLYHFMSQFIDKSQQ
 LGLTHYIFMNELKINLLDFREQMDNLEAIGLIKTFVRHEEKYSHFVYELIQPPTAYQFFN
 DPMLSVFLFSEVDKKRYQALKSYFEKDEKDLISKYQQTRKFTEVFNVPKKVNVSQINLK
 5 QIKHYDGLDLSNETDFEMLRQMLNHHFISNEIIDKEAKNLIIQLATLYGITEDGMKNVI
 LSSITSAQQLSFEEMRKARTYYLIEHDNQLPKLEHQTNKINDEKKDRQAEDTTNDWLQL
 LDETSPIDMLASWSDSEPTQSQKSMIEELINREKMNFGVINILLQFVMLKEDMKLPKSYI
 FEIASNWKKIGISNAKQAYEYALQVNQPKNYETHSNDKRQNNRGRQNQFLSKEKTPKWLO
 NRDDQEEENKEINDDTLEEDRQAFLEKLNQKWKEEDN*

Sequence 1191

Contig_0569_pos_10279_11178,

is similar to (with p-value 6.0e-63)

>sp:sp|P06567|DNAI_BACSU PRIMOSOMAL PROTEIN DNAI. >pir:pir|B

15 24720|IQBS44 dnaA protein homolog, 44K - Bacillus subtilis >

gp:gp|AF008220|AF008220_192 Bacillus subtilis rrnB-dnaB geno

mic region. NID: g2293135. >gp:gp|X04963|BSDNAB_1 Bacillus s

ubtilis dnaB gene for initiation of chromosomal replication.

NID: g39880. >gp:gp|Z99118|BSUB0015_163 Bacillus subtilis c

20 omplete genome (section 15 of 21): from 2795131 to 3013540.

NID: g2635200. >gp:gp|Z75208|BSZ75208_2 B.subtilis genomic s

equence 89009bp. NID: g1769994.

atgggcgattctcaaaatctagataaaacgtatacaaaaaataaaacaaatgtaatcaat

gatactgacgttaaacattttcttgagaaaaatcgtagtaataataactaatgagatgata

25 gacgaagatttaaatgttcttcaagagtataaaagatcaaaaaagtttatgatggacat

cgctatgatgattgtccgaattttgtaaaaggacatgttctgaactatatattgaaaat

gaaagaatcaaaattagatatctaccttgcccggtgtaaaattaaacatgatgaggaacga

tttgattcacaaacttattacatctcaccatattgcaaagagatacacttcatgcaaagctc

aaagatatatttatgaataatcgagagagacttgatgtagcaatggcagctgatcaaatc

30 tgtacagcaattactaacgatgaaaaagtaaaggggttatatttatatgggtccttttggt

acaggaaaaatcattcatattgggtgctattgcaaatcaacttaaatcgcaaaagatttca

tcaacaattgttatattaccagaatttatttcgcactttaaaaggtggctttaagacggt

agttttgagaaaaaattacaacgtgtgcgagaagctaataattttgatgttagatgatatt

ggcgcagaagaagtcacaccgtgggtaagagatgaagtgattgggtcctttattacattat

35 agaatggtacatgaacttcctacatttttagttctaactttaattatagtgagcttgag

catcatctttcaataactagagatggcactgaaaagactaaagcagcacgaattattgaa

agaattaagactttatcgacaccttattatttgactggtaaaaattttagaacaattga

Sequence 1192

MGDSQNLDKRIQIKQNVINDTDVKHFLEKNRSNITNEMIDEDLNLVQLEYKDQKQVYDGH

RYDDCPNFVKGHVPELYIENERIKIRYLPCKIKHDEERFDSQLITSHMQRDTLHAKL

KDIYMNNRERLDVMAADQICTAITNDEKVKGLYLYGPFGTGKSFILGAIANQLKSQKIS

STIVYLPEFIRTLKGGFKDGSFEKKLQVRANILMLDDIGAEVTPWVRDEVIGPLLHY

45 RMVHELPTFFSSNFYSELEHLSITRDGTEKTKAARIIERIKTLSTPYLTKNFRNN*

Sequence 1193

Contig_0569_pos_0_389,

is similar to (with p-value 1.0e-19)

>gp:gp|U56999|TPU56999_1 Treponema pallidum methyl-accepting

chemotaxis protein (mcp-1) gene, complete cds, and potentia

l regulatory molecule (pfoS/R) gene, partial cds. NID: g1354

774.

55 atgtcgacgataaaaaatatacgatggaccaaaggattttgttttttagagtgttatcaggg

gtagcaattggaatagtagccggactcggttccaaatgcaattttgggagaaatttttaaa

tacttttatgcaatatcatcctattttcaaaactttattaggggtcgttcaagccatccaa

tttacagtgccagcgcttattggagcattgatagctatgaagttcaatatgacaccttta

gcaatagctgtagtagcaagtgccctcatatgttggtagtggtgcagctcaattttaacaa

ggtagcatgggtaattgcgggaattggggatttaattaatacgaatgtaactgcatccatc
gccgtgttacttattttattaatagaaga

Sequence 1194

5 MSTIKNIDGPKDFVFRVLSGVAIGIVAGLVPNAILGEIFKYFMQYHPIFKTLLGVVQAIQ
FTVPALIGALIAMKFNMTPLAIAVVASASYVGSAAQFKQGTWVIAGIGDLINTMLTASI
AVLLILLIEX

Sequence 1195

10 Contig_0570_pos_734_1495,
is similar to (with p-value 4.0e-40)
>sp:sp|P54717|YFIA_BACSU_HYPOTHETICAL_29.3_KD_PROTEIN_IN_GLV
G-GLVBC_INTERGENIC_REGION. >gp:gp|Z99108|BSUB0005_88 Bacillu
s subtilis complete genome (section 5 of 21): from 802821 to
15 1011250. NID: g2633055. >gp:gp|D50543|D50543_2 Bacillus sub
tilis DNA for 76-degree region, complete cds. NID: gl486240.
atgatttttagatgaacgtgttaaactctaatttcgatcaattaaatgataatgatatacaa
attgcacattatgttaatacacatatagatgtttgcaaaaatatgaaaatacaagattta
gcctcacagacacatgcttcaaatgctacgattcatcgcttcactcgtaaaactagggtttt
20 gatgggtatagtgaactttaaactccttttttaaaatttgaagatagtaagaatcatcaactt
ccttctgattctatggagcaatttaaacagaattgaaaatacattcaactatttagaa
cgtattgattatcggtttattaactcacaaaatgcatcatgctacaacaatatacttatat
ggtactggacgtgcacagatgaatgtcgctgaagaagcacaaactatactgttgactatg
cataaaaatatttatattgttacatgatgttcatgaactaaagatgggtgttaaacagaaca
25 attccagaagatttgtttttcatcatttcactttctggcgaaacacatcaacttaagaa
gtcacacaattgcttcaactgagacaaaaatattttatttccgtaacaacaatgaaagac
aatacattggcacaacaagctgattacaatgtctatgtttcaagcaataccttctattta
aacgatggtagtatttccagttttattagctatcacattttctttgaaacactacta
agaaaatataacgaatataaagagaatcatgaattaacatag

Sequence 1196

35 MILDERVNSNFDQLNDNDIQIAHYVNTHIDVCKNMKIQDLASQTHASNATIHFRTRKLG
DGYSDFKSLKFEDSKNHQLPSDSMEQFKQEIENTFNLYERIDYRLLTHKMHHATTIYLY
GTGRAQMNVAEEAQRILLTMHKNIILLHDVHELKMLVNLKTIPEDLFFIISLSGETHQLKE
VTQLQLRQKYFISVTMKNLTLAQQADYNVYVSSNTFYLNQDGTDYSSFISYHIFFETLL
RKYNEYKENHELT*

Sequence 1197

40 Contig_0570_pos_2036_3049,
putative peptide of unknown function
atggaacgattttgtgtgtgaaatcaaattaactatattcaaatgaatccgttagaagcc
aaatttaaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagcctt
gcttgttttaggaccgacgcttaaacaaacagacagcttacctatacatgagttaatatc
tttgaaattaagagaacgcgtccgttttcatctagaatcgagaatgaacaaaaatcgactt
45 aaatttcagatccttgaattactccatcaaacattccctgggttagaaagattgtttagt
agtcgatattcaatcattgcactcaacatcgagaaatctttactcatccagacatgggtt
cttgatatcgacaaggaggtactgattacacatatattcaattctacagataagggaatg
tcaatggataaagctacaaaatattgcacttcaattaagggtgattgctcaagaaagctat
cctagtgtcgatagacattcctttctagtgcgaaaattacgcttacttattcaacaatta
50 aaacaatctattcatcatctcaacaatttagatgatgccatgattcaattagcacacaa
ctcgattatttgaaaatattcattcgatacctgggtattggttaagctaagcacagctatg
attattggggagattgggtgatattaagcgatttaaatcaataaacaactcaatgctttt
gttggcattgatatacaacgatataatcaggtcatacacactgtagagataccatcaac
aagcgtgggtaataaaaaagcgagaaaacttttattttgggtgattatgaatataataaga
55 gggcagcatcattatgacaatcatgtcgtcgattattactacaaactaagaaagcagcct
aatgagaaacctcataagactgccatcattgctgtataaatcgattattaaaaacaatt
cattatcttgaatgaatcataaattgtacgattatcaaatgtcaccacattag

Sequence 1198

MERFCCVNQINYIQMNPLEAKFKTSALRSWKTDQADAHKLACLGPTLKQTDLSLPIHELIF
 FELRERVRFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSI IALNIAEIFTHPDMV
 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPSVDRHSFLVEKLRLLLIQQL
 KQSIHHLKQLDDAMITQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFKSNKQLNAF
 5 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLFWVIMNIIRGQHHDYDNHVVDYKYLRKQP
 NEKPHKTAIIACINRLKTIHYLVMNHKLYDYQMSPH*

Sequence 1199

Contig_0571_pos_1443_2774,

10 is similar to (with p-value 0.0e+00)

>sp:sp|P13375|G6PA_BACST GLUCOSE-6-PHOSPHATE ISOMERASE A (GP
 I A) (EC 5.3.1.9) (PHOSPHOGLUCOSE ISOMERASE A). >pir:pir|S15
 936|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - Ba
 cillus stearothermophilus >gp:gp|X16639|BSPGIA_1 Bacillus st
 15 earothermophilus pgIA gene for phosphoglucoisomerase isoenzy
 me A (EC 5.3.1.9). NID: g40045.

atgactcacattcaattagactatggcaaaacttttagaattttttgataagcatgaacta
 gatcagcaaaaaggatattgttaaaactatccatcaaactattcataaaaggtagaggagca
 ggtaatgacttttttaggttggttagatttacctgttgattatgataaagaagaattttct
 20 agaatcgctcgaagcatctaaacgtatcaaatacaattccgatgtacttgttggtatcggt
 attggaggttcatacttaggtgcacgtgctgcaatcgagatgcttacatcttcatttaga
 acaaatacggaataccctgaaattgtattgttaggtaatacatttatcctcaagttataca
 aaagaattacttgattatttacaaggaaaagattttcagtttaacgttatttcaaaatca
 ggtactacgacagaaccagcagttgcatttagattatttaacaattggttgaagaaaaa
 25 tatggaaaagatgaagctaagaaacgtatttttgcaacgacagataaatctaaaggtagca
 cttaaaacaattagcagacaatgagggttatgagacgtttgttgtagctgatgtgga
 ggtagcttattctgttcttacagctgtaggattactaccaattgcaactgcaggtatcaat
 attgaatcaatcatgattggtgctgcaaggcacgtgaagagttatcttctgatgattta
 gatcaaaatatcgcatatcaatatgcaactattcgaaatattttatcacgaaaagggtat
 30 actactgaaatgttaattaattacgaacccctctatgcagttatttcaacgaatggtgga
 caattatacgggtgaatcagaagggaagatttcaaaaggattttatccatcaagtgcgaat
 tacacaactgatttacattccttaggacaatatgttcaagagggccgctgttcttattc
 gagacagtggttaagggtcaaccatccaaaacatgatataaaaattgaagaggatgcagat
 gatttagacggactgaactatcttgctggcaaatcaatcgatgaagtgaataactaaagca
 35 tttgaaggtagacttctgcacataccgatggtggcgttccaaatatcggttgtaaatatt
 cctcagtttagatgaagaaacatttggatatgttggtttatttctttgaattagcttgtgca
 atgagtgatatacaatttaggtgtaataccatttaatacaacctggagttgaagcctataaa
 caaatatggttgcgctattaggttaaaccaggctttgaagataagaaaaagaattagaa
 aatcgtttataa

40

Sequence 1200

MTHIQLDYGKTLEFFDKHELDQKQDIVKTIHQTIHKGTGAGNDFLGWLDLPVDYDKEEFS
 RIVEASKRIKSNSDVLVIGIGSSYLGAARAAIEMLTSSFRNTTEYPEIVFVGNHLSSTYT
 KELLDYLGKDFSVNVISKSGTTTEPAVAFRLFKQLVEEKYKDEAKKRIFATTDKSKGA
 45 LKQLADNEGYESFVVPDDVGGRYSVLTAVGLLPATAGINIESIMIGAAKAREELSSDDL
 QONIAYQYATIRNILYSKGYTTEMLINYEPSMQYFNEWKQLYGESEKDFKGIYPSSAN
 YTTDLHSLGQYVQEGRRFLFETVVKVNHKPKHDIKIEEDADDLDGLNYLAGKSIDVNTKA
 FEGTLLAHTDGGVFNIVVNIQQLDEETFGYVVYFFELACAMSGYQLGVNPFNPQPGVEAYK
 QNMFALLGKPGFEDKKKELENRL*

50

Sequence 1201

Contig_0571_pos_0_1100,

55 is similar to (with p-value 0.0e+00)

>sp:sp|P50986|ASSY_STRCL ARGININOSUCCINATE SYNTHASE (EC 6.3.
 4.5) (CITRULLINE--ASPARTATE LIGASE). >pir:pir|S57659|S57659
 argininosuccinate synthase (EC 6.3.4.5) - Streptomyces clavu
 ligerus >gp:gp|Z49111|SCARGGH_1 S.clavuligerus argG gene and
 argH gene (partial). NID: g886905.
 atgaaagataaaatcgtttttagcatattcaggtggttttagatacaagcgttgcaagttcaa

tggccttattgataaaggatatgatgtagttgcttgttgccttgacgtaggcgaaggcaaa
 gatttagacggttgtatatcaaaaagcttttagatatgggtgcagtcgaatgtcatattatt
 gatgcaactaaagaatttagtgatgattatgtaagttatgctattaaaggaaatttaag
 tatgaaaaatgcatatcctctagtttcagcattatcacgtccactcatcgcaaaaaaactg
 5 gttgaaattgctgaaaaaacaattctatttggtattgcgcatggatgtactggtaaagg
 aatgatcaagtacgtttcgaagtggcaatcaaagctttaaatcctaagttaaaagcattt
 gcacctgttcgtgaatgggcttgagcagagaagaagaattgattacgcaatcaaacat
 aatattcctgtttcaatcaattatgactcgccatactcaattgaccaaacttatggggg
 agagctaataatgtgggtattttagaagatccgtatgccgcacctccggaagatgcattt
 10 gatttaactacacctttagaagaactccagacaatgcagacgaaattatccttacattt
 aaacaagggtattccagtagcaagttgatggcaagattatcaattagatgaccttattctt
 tacttgaatcaacttgctggcaaacacggtattggtagaatcgatcatgttgaaaacaga
 atggctcgggataaaaatcgagagagattttatgaaacacctgggtgcggaagttattttaaa
 gcacacaaaagcactagaaacaattacattaactaaagacgtagcgcactttaagcctgtc
 15 attgaaaaacaattttcagaacaaatatacaatgggttgggttctcgccattaacagat
 agtttaaaactctttatcgatagtactcaacaatatgttgagggagatgtgagaattaaa
 ttatttaagggaacgctattgtcaatggcagacaatctccttacactttatacagatgaa
 aaattagctacttatacGGA

20 Sequence 1202

MKDKIVLAYSGGLDTSVAVQWLIDKGYDVVACCLDVGEGKDLVVYQKALDMGAVECHII
 DATKEFDDYVSYAIKGNLMYENAYPLVSALSRPLIAKKLVEIAEKTNSIGIAHGCTGKG
 NDQVRFEVAIKALNPCLKAFAPVREAWWSREEEIDYAIKHNIIPVSINYSIDQNLWG
 RANECGILEDPYAAPPEDAFDLTTPLEETPDNADEIILTFKQGI PVQVDGKDYQLDDLIL
 25 YLNQLAGKHGIGRIDHVENRMVGIKSREIYETPGAEVILKAHKALETITLTKDVAHF KP
 IEKQFSEQIYNGLWFSPLTDSLKLFIDSTQQYVEGDVRIKLFKGNAIVNGRQSPYTLYDE
 KLATYTX

Sequence 1203

30 Contig_0572_pos_1688_2902,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P39141|NUPC_BACSU PYRIMIDINE NUCLEOSIDE TRANSPORT PRO
 TEIN. >gp:gp|D45912|D45912_25 Bacillus subtilis genome seque
 nce between the iol and hut operon, partial and complete cds
 35 . NID: gl408482.

atgcatattgtgattgggatattaggaatcattttcttttagcactcgagttttattt
 agttcagacagaaaaaatattcgctggcgatattgttgattgctattagtaattcaacta
 atatttgcatttatattacttaaaactaatttgggaatttcagttattgggagatttca
 gatgggttttaattatttatttagctaaagcagcggtcggtgtcaattttgtatttgggtggc
 40 tttaaatttattgatcctaaacaaccaccattcttcttttagcgttttgttacctattgtt
 tttatttcagcattgattggtatattacaatatacacgaatacttccactaattattaac
 ttactgggctttttaatttcaaaaattaatggaatgggcccgtttagaatcttacaatgcg
 gtcgcggcagcaattctaggacaatctgaagtctttatctcattaaaaaacaattacct
 tacatacctaacaacgcttatatacattaactgcttcagcgatgtcaacggtatcagca
 45 tcaattatagcgcttattttacacttattgaacaaaatatgttggttactgcagtagtg
 cttaaacttgtttgggtgtttatcattgcatctatcattaatccttataaagtcaatgag
 gaagacgacaaattattaattgatgagaacgaaacaaaaaacaatctttctttgaaatg
 ctggggagtatatactagatggatttaagtagcagttattgtaggcgctatgctgata
 gggtatattgcaattattgctttattaaatggaatgggtgagtggaatcttaagctttatg
 50 tctgggtggtgctattcaatggaacttccaaacgcttattggatttatttttgcaccttc
 gctttcctaactggaataccgtggcaagatgcagttcaatctgggtcagtaattggctaca
 aaattactactaaatgaattttagcaatgcaagatttaggtaaagcgactggattatcg
 gaacatgctaaggaattacctctgtcttcttagtatcattcgcaaaactttagtcaatt
 ggtattatttcaggagctattaaatcattgaatgatgaaaaagggtgacgttgttgctcgt
 55 ttcggaataaaaattattttggtgcaacacttgtttcgctttatatcagcggtattgca
 ggattctttatctaa

Sequence 1204

MHIVIGILGIIFFLALAVLFSSDRKNIRWRYVGLLLVIQLIFAFILLKTNLGISVIGSIS

DGFNYLLAKAAVGVNFVFGGFKFIDPKQPPFFFSVLLPIVFISALIGILQYTRILPLIIN
 LLGFLISKINGMGRLESYNAAVAAAILGQSEVFISLKKQLPYIPKQRLYTLTASAMSTVSA
 SIIGAYFTLIEPKYVVTAVVLNLFGGFIIASIINPYKVNEEDDKLLIDENETKKQSFFEM
 LGEYILDGFKVAVIVGAMLIGYIAIALLNGMVSGILSFMSSGAIQWNFQTLIGFIFAPF
 5 AFLTGIPWQDAVQSGSVMATKLLSNEFVAMQDLGKATGLSEHAKGITSVFLVSFANFSSI
 GIISGAIKSLNDEKGDVVARFGIKLLFGATLVSFISAAIAGFFI*

Sequence 1205

Contig_0572_pos_6336_6677,
 10 is similar to (with p-value 1.0e-28)
 >gp:gp|U35659|SBU35659_1 Streptococcus bovis malic enzyme ge
 ne, complete cds. NID: g1006838.
 gtgtcagttgtatctagagctattggcacaccattaatacccgcaaaacttttgaacagt
 gctgccttttccttcattactggaatacttgcttctgctccaatattccctaaaccgaga
 15 acagcagttccatctgtttacaacagcaactgtatttcctttaatagtgtactcatatact
 tttcttgaatcttcatggatttctttacaaggttctgcaacgccaggtgagtagtctagg
 ctttaattgttgcttatttgtcactttaacatttggtgtaatttctagtttaccttggttc
 tctctatgcatttctaagcgtcatctcttaagacatttaa

20 Sequence 1206

VSVVSRAIGTPLIPAKLLNSAAPSITGILASAPIFPKPRTAVPSVTTATVFPLIVYSYT
 FLESSWISLQGSATPGEYARLNCCLFVTLTFGVISLPLWFSLCISKASSLKDI*

Sequence 1207

25 Contig_0572_pos_11556_11170,
 putative peptide of unknown function
 atgcaattattaggtagacaacctcaaatgggtgaaacagtcattgatcaaattgccaaa
 catatttcaatacatcaacaaggaataaatgtcgacgtatctccactcattacaaatcat
 tacggtacactaagtaaagcagttttgttgaattattgaggaaacgattcggcagcaa
 30 atgagaaagtataaaaaaggtaatgtcatgatagaaagtatgagtattatctatattaag
 actgtaccaattgaatctactattgaagtacattatgaaatgttagatgttggccgatat
 tttgctaaattagaagttactatgattaataatgggtgaaaaagttgctaattgcattagta
 atttgtcaaattgttgatgggttttaa

35 Sequence 1208

MQLLGRQPQIGETVNDQIAKHISIHQQGINVDVSPLITNHYGTLSKAVFVGIIETIRHE
 MRKYKKGNMIESMSIIYIKTVPIESTIEVHYEMLDVGRYFAKLEVMTINNGEKVANALV
 ICQMFDDGF*

40 Sequence 1209

Contig_0572_pos_11136_10198,
 is similar to (with p-value 4.0e-31)
 >sp:sp|P22746|MGP_MYCGE MGP PROTEIN. >pir:pir|A64221|A6422
 1 MgPa operon 29K protein homolog MG190 - Mycoplasma genital
 45 ium (SGC3) >gp:gp|U39698|U39698_5 Mycoplasma genitalium sect
 ion 20 of 51 of the complete genome. NID: g3844782.
 atggaagtaaaaatgaatgaaataatggaagcattagaacaaagtgaattaattattatt
 cacagacatctaagaccagatccagacgcataatggttcacaattaggtttgaaatattac
 ttacaaaagaagtttccaaacaaacaaatttatgctgtaggagctaataagattccttg
 50 aaatttataggtttgatggacgaaattgacaatgatataacaagaaagcgactgtagtt
 gtatgtgatacggcaaatgcccacgaatagatgaccaacggttatgatacaggtaccaa
 cttttgaaaattgatcatcatcctgctactgatcagtaggagatattaactatgttaat
 accaaagcttctccactagtgaataatttacgaattcatttcacatttcaatgatgaa
 catatcattgatgaacaagttgctagagattatattcttggcatcgttgggtgatactgga
 55 cgttttttatttaataataacaacgccaagaacaaatgcaaattgctggaaaattacttaca
 tatccttttgatcacaaacgaattaaacaaaatgtctgaaaaggatccaaaactatta
 ccatttcaaggatataattgcaaaattttgatttaaatgataaaggattttgcaaagtt
 aaaataactaaagacatacttgaaaatttcaaatacaacctaataagcgtctttattt
 gtaataacaatcgagatattcgaggattaaaaataggatgttggcgttgatgaagga

gatcaaattagatgtcggttgcgttctaaaggtcatattattattaatgatgtcgctaata
acatttgggtgggtggacatccaaatgcatctggagtttcagtaaatagttgggagcaa
ttcgagcaactcgccgaagctttaaacgacaagttataa

5 Sequence 1210

MEVKMNEIMEALEQSELI I IHRHLRPDPDAYGSQLGLKYLLQKKFPNKQIYAVGANEDSL
KFIGLMDEIDNDIYKKATVVVCDTANAPRIDDRYDTGKLLKIDHHPATDQYGDINYN
TKASSTSEIIYEFISHFNDEHIIDEQVARVLYLGIVGDTGRFLFNNTTPRTMQIAGKLLT
YPFHDHNQELNMSEKDPKLLPFQGYILQNFDLNDKGFCVKITKDILEKFQIQPNEASLF
10 VNTIADIRGLKIWMFGVDEGDQIRCLRSKGHIINDVANTFGGGGHPNASGVSVNSWEQ
FEQLAEALNDKL*

Sequence 1211

Contig_0572_pos_10174_6977,
15 is similar to (with p-value 0.0e+00)
>sp:sp|P14567|DP3A_SALTY DNA POLYMERASE III, ALPHA CHAIN (EC
2.7.7.7). >pir:pir|A45915|A45915 DNA-directed DNA polymerase
(EC 2.7.7.7) III alpha chain - Salmonella typhimurium >gp:
gp|M29701|STYDNAE_1 S.typhimurium polymerase III polymerase
20 subunit gene, complete cds. NID: g153951.
atggtagcacatttaaatattcacttcttttgacctgtagattctagtttaagaatt
gatgcattaatagataaagctaaaaaagaaggatatcggtgcgcttgcaataaccgatata
aatgtattgtacggttatccaaagttttatgatgcttgtattgcagctcacatacatcca
atctttgggtatgactatatatttaacggatggtctctatactattgaaacggttggttta
25 gcaaaaaataatcaagcagaattcaagtcattatatacaactttctctgctataatgatgaga
aataaagaagaagtgccaattgaatggctaaaaagatacgacgaacatttaattatcata
tttaaaagaggtgagttgtctcataagcaagttattgatgcttttgaaggttaagaaagaa
ttatatttaaatcacaaatagtaataatacattgactggcaaacgtgtatggatgcaatct
gcaagatacttaaatgaagatgatgctgaaaccattccagcgttacatgccataagagat
30 aatactaagtttagatttaatacatgagaaagaacacttgatgaacattttcctagtata
gaagaacttcaaacactaaatcttagtgaaagatgattactaacgcgaatgaaattgaa
gaattatgccaaagcagaattgcataccatcaatccctggtgccacaatttgtgacacct
aatggtgaaacttcgaaagattatctttggacgatacttatacataggttacgagaatgg
gaacttaatgataaaacttatttcaatcggttgaaacatgaatataaaattattactgat
35 atgggtttcgaggattattttcttatttgtaagtgatttgattcattttgctaaacacat
gaagtgatgggttgggccaggtcggtggttcacagcaggtcattagtaagttatttatta
ggtattactactatagaccggttaaaatataatcttttatttgaaagattttcttaactct
gaacgcgtaactatgccagatattgatattgattttgaagacacgagacgtgaaaaagta
attaatgtatgtacaagataaaatattggtgaacatcatgtatcaggatttgtgacatttggg
40 catctgttagctcgtgctgttgctagagatgttaggaagaataatgggatttgatgaaacg
agtttaaatgagatttcaaaacttattccacataaattaggtataactcttgaaagaagca
taccaaaagccagagtttaaaagcatttgttcacgtaatcatagaaatgaacgttggttt
gaagtgagtaaaaaagttagagggattaccaagacatacgtctacgcatgctgcaggatc
attatcaatgatcaaccattattcaaatttgccccattacaactggtgatacaggatta
45 ttaacgcagtggtgactatgacagaagcggaaacgtataggattatataaaattgatttcttg
ggattacgcaatctatcaattattcatcaattattttacaagttaaaaaggatttaaat
ataaatattgatatagaagctataccttatgatgataaaaaagtttttgatttattatca
aacggtgacactacaggtatatttcaattggaatcagacggtgttagaagcgtattaaaa
agattgcaaccggaacattttgaagatatcgtagctgtcacatcatttatagaccagga
50 ccaatggaagaaataccaacttatataaccgtagacataatcctaaccaatttgcttat
ttacatccagatttagaaccaatcttaaaaaacacatatggtgttatcatttatcaagaa
caaataatgctaataagcaagtcaggtgctggttttagttatggtgaagcagatatttta
agaagggcaatgagtataaaagaatcggtgcaatcttagaaaagtgaagcgtcaacatttcat
gatggtgcaaaaaataacggttacgatgaacagataagtaagcaaatttttgatttaata
55 cttaagtttgagattatgggttcccacgtgcccatgctgttagttactcaaaaattgca
tacattatgagctatttaaaagtcactatcctcattattttatgcaaatatcttgagt
aatgtaataggaagtgaaaaaagactgcagctatgattgacgaagctaagcaccaaaaga
attagcatcttgctcccaatattaatcaaagtcatttggtattataaggcaagtaataaa
ggaatatatctgtcttttaggtacaattaaaggaattggatatcaaagcgttaaatatt

attgatgaacgtcagcagaatggaccttatagagatttctttgatttttcaagacgtata
 ccaaaaagggtgaaaaatagaaaattacttgagtccttcttagtaggcgcattcgac
 acttttggcaaaactagagcgacattattacaagcaattgatcaagtattagatttgaat
 tctgatgttgagcaagatgaaatgcttttcgatcttttaactcctaacaatcgatgaa
 5 gaaaaagaggaactacctgatcaattattaagtgattatgaaaaagaatacctaggattc
 tatattagtaaacatccagttgaaaagaaatttgaagaagaacaatatttaggcataatt
 caattgtctaattggaagtcactaccaacctatacttggtcaatttgaccatatcaacaa
 ataagaacgaagaatggtcaaaatatggcatttgaacgatgaatgatggaagaacgatg
 atggatggagtgattttcccagataagtttaaaaaatacgaaacttctatttcaaaggaa
 10 cagatgtatatcgattaggtaaatttgaaaagcgtaaccaacaatgcaacttatcatc
 aatcaactttttgaagttgaagcgatgagcaaaacaaattgtctaattcgaaaaagtt
 attttacgtaatgtaacacatctagaaccacaatttgaacattcaaaagtagaatcta
 gaacaacatgcattaaatatttatggttttgacgaaagtgcaataagatgacaatgttg
 ggacaaattgaacgtcaacgtcaaaattttgatctattaatacaaaacttattcgccagct
 15 gatattagattttatttaa

Sequence 1212

MVAHLNIHTSFDDLSSLRIDALIDKAKKEGYRALAITDTNVLYGYPKFYDACIAAHHP
 IFGMTIYLTDLGTYTETVVLAKNNQGLKSLEYQLSSAIMMRNKEEVPIEWLKYDEHLIII
 20 FKEAELSHKQVIDAFEGKKELYLNHNSNNTLTGKRVWMSARYLNEDDAETIPALHAIRD
 NTKLDLIHEKETLDEHFPSIEELQTLNLSSEDMITNANEIEELCOAEIAYHQSLLPQFVTP
 NGETSKDYLTILHRLREWELNDKTYFNRLKHEYKIITDMGFEDYFLIVSDLIHFAKTH
 EVMVGPGRGSSAGSLVSYLLGITTIDPLKYNLLFERFLNPERVTMPDIDIDFEDTRREKV
 IKYVQDKYGEHHVSGIVTFGHLARAVARDVGRIMGFDETSLEISKLI PHKLGITLLEEA
 25 YQKPEFAFVHRNHRNERWFEVSKKLEGLPRHTSTHAAGIIINDQPLFKFAPLTTGDTGL
 LTQWTMTAERIGLLKIDFLGLRNLSIIHQIILQVKKDLNINIDIEAIPYDDKKVFDLLS
 NGDTTGIFQLESDGVRSLKRLQPEHFEDIVAVTSLYRPGPMEEIPTYITRRHNPQFAY
 LHPDLEPIKNTYGVIIYQEQIMLIASQVAGFSYGEADILRRAMSCKKNRAILESERQHFI
 DGAKNNGYDEQISKQIFDLILKFADYGFPRHAVSYSKIAYIMSULKVHYPHYFYANILS
 30 NIVIGSEKKTAAMIDEAKHQIRISILPPNINQSHWYKASNKGIYLSLGTIKGIGYQSVKLI
 IDERQNGPYRDFGFSRRI PKRVKNRKLLESILVGAFTFGKTRATLLQAI DQVLDLN
 SDVEQDEMLFDLLTPKQSYEEKEELPDQLLSDEYEKEYLGFYISKHPVEKKFEKKQYLGIF
 QLSNGSHYQFILVQFDHIKQIRTKNGQNMAFVTMNDGRTMMDGVIFPDKFKKYETISKE
 QMYIVLGKFEKRNQMQQLINQLFEVEAYEQTKLSNSKKVILRNVTLEPQFEHKSKEVN
 35 EQHALNIYGFDESANKMTMLGQIERQRQNFLLIQTYSADIRFI*

Sequence 1213

Contig_0572_pos_6673_6230,
 is similar to (with p-value 6.0e-41)
 40 >gp:gp|U35659|SBU35659_1 Streptococcus bovis malic enzyme ge
 ne, complete cds. NID: g1006838.
 atgtctttaagagatgacgcttttagaaatgcatagagagaaccaaggtaaactagaaatt
 acaccaaatgttaaagtgacaaataagcaacaattaagcctagcatactcacctggcggtt
 gcagaacctgttaaagaatccatgaagattcaagaaaagtatatgagtacactatttaa
 45 ggaataacagttgctgttgaacagatggaactgctgttctcggtttagggaatattgga
 gcagaagcaagtattccagtaattggaaggaaaggcagcactgttcaaaagttttgcgggt
 attaatggtgtgccaatagctctagatacaactgacactcaagaaatcataaaacagta
 aaacttattgcacaaactatggtggaattaatcttgaagatatatcagctccatttta
 tattgggttcaaaacatggtattga
 50

Sequence 1214

MSLRDDALEMHRENOGKLEITPNVKVTNKQQLSLAYSPGVAEPCKEIHEDSRKVYEYTIK
 GNTVAVVTDGTAVLGLGNIGAEASIPVMEGKAALFKSFAGINGVPIALDITDTQEIITKV
 KLIAPNYGGINLEDISAPILYWFKTWY*
 55

Sequence 1215

Contig_0572_pos_6222_5281,
 is similar to (with p-value 4.0e-48)
 >gp:gp|AF068902|AF068902_4 Streptococcus pneumoniae D-glutam

ic acid adding enzyme MurD (murD), undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase (murG), cell division protein DivIB (divIB), orotidine-5'-decarboxylase PyrF (pyrF), and orotate phosphoribosyltransferase PyrE (pyrE) genes, complete cds; and unknown genes. NID: g4009477.

5 atgatagagtcacaaactccctgatattcaatattatccaatatcaagcggtaaattacgtagcttatctatcttttgaaaatgcaaaagatgtctttaaagtttgaaaggaatttagatgcacgtaaaaatacttaaaaaacaaaaccagacttactttttcaaaagggtggtttgttagtgttccggtagttatagccgcacgttctttaaaaattccaactatcatcacgaatcagatttaactcctggattagctaataaaatttctttaaaatttgctaagaaaatatacacaacctttgaagatacacttacatatcttccaaaagataaagctgattttgttgggctactgtacgtgaggacttaaaacaagggataaaagaaaggagatcaactcactgattttgataaaaataaaaaagtgttattagtcatgggaggaagtttaggtagtaaaaaacttaataat

10 atcattcgtcaaaatattgaggcacttctccacgattatcaaattatacacttaactgga

15 aaaggacttggtgatgactcaatcaataaaaaagggttatgttcaatttgaaattgttaaa

gacgacttaactgatttattagcaatcactgatactgttgtaagtcgtgcagggttctaacgcaatttatgaatttttaacgctacgtataccgatgttactcatccccttaggacttgatcaatcaagaggagatcaaaattgataatgctaaaaactttgaatcaagggttatggtcgcatattcctgaagatcaacttacagaagtttaacttattgcaagaattaaatgatattgaa

20 ttacatcgtgaatctattattaacaaatggaaacatatcaagagagttacacgaaagaa

gatttatttgataaaattattcatgatgcattaaacaagtag

Sequence 1216

MIESQLPDIQYYPISSGKLRRYLSFENAKDVFKVLKGILDARKILKKQKPDLLFSKGGFV

25 SVPVIAARSLKIPTIIHESDLTPGLANKISLKFARKIYTTFEDTLTYLPKDKADFGAT

VREDLKQGNKERGYQLTDFDKNKKVLLVMGGSLSKSKLNNIIRQNI EALLHDYQIIHLTG

KGLVDDSIKKGYVQFEFVKDDLTDLLAITDTVVSRAISNAIYEFLLRIPMLLIPLGLD

QSRGDQIDNAKNFESKGYGRHIPEDQLTEVNLLQELNDIELHRESIIKQMETYQESYTKEL

30 DLFDKIIHDALNK*

Sequence 1217

Contig_0572_pos_5268_4654,
is similar to (with p-value 3.0e-21)
>pir:pir|S32217|S32217 hypothetical protein 2 - Bacillus meg

35 aterium >gp:gp|Z21972|BMCTP450A_3 B.megaterium cytochrome P4

50meg, ORF1 and ORF2 genes. NID: g288298.

atgaatcgatggaaacgcatttcattgcttattgtttttacacttatttttggtataata

gctttttttcatgaatcaaggcttggaatggatagataacgaagtatatgaatttatt

40 tttcatctgaagtttcattaccacatctattatgttaggtgtaacaaaattgggtgaa

gtttgggcaatggttgcgctatccttattattagttgcttaccctatgctaaaacgcttc

aagattgagacattattctttgtaatagtaatgagcttatctagtacactcaatccacta

ttaaagaatatctttgataggaacgtccaacattattgcgtttaattgacatttcaggc

tttagttttccaagcggctcatgctatgggctcaacttcattctttggaagcgctatatat

45 gtaataaacgcgtcatgattcgggtatctctaaaggcgtgtaatcggtttatgcgcactt

ttcatttttattaatatcaacttctagagtgtatctaggcggttcattaccctacagatatt

attgccggcattattggtggtgtattctgccttttactcagttacttttactacctaata

cagttaatagcttag

Sequence 1218

MNRWKISLLIVFTLIFGIIAFFHESRLGKWIDNEVEFIYSSSEFITSIMLGVTIKIGE

50 VWAMVALSLLL VAYLMLKRFKIETLFFVIVMSLSSTLNPLLNIFDRERPTLLRLIDISG

FSFPGHAMGSTSFFGSAIYVINRHDSGISKGVLLIGLCALFILLISTSRVYLGVHYPTDI

IAGIIGGVFCLLLSTLLLPKQLIA*

Sequence 1219

Contig_0572_pos_4081_3068,
putative peptide of unknown function

atggaacgattttgttggtgtaaatacaatcaactatattcaaatgaatccggttagaagcc

aaatttaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagctt

gcttggttaggaccgacgcttaaacaacagacaacttacctatacatgagttaatattc
 tttgaattaagagaacgcgctccgttttcatctagaaatcgagaatgaacaaaatcgactt
 aaatttcagatcccttgaattactccatcaaaccattccctgggttagaaagattatttagt
 agtcgatattcaatcattgcaactcaacatcgagaaatctttactcatccagacatgggt
 5 cttgatatcgacaaggaggtactgattacacatatattcaattctacagataagggaatg
 tcaatggataaagctacaaaatatgcacttcaattaagggtgattgctcaagaaagctat
 cctaattgctgatagacattcctttctagtcgaaaaattacgcttacttattcaacaatta
 aaacaatctattcatctcaacaattagatgatgccatgattcaattagcacaacaa
 ctcgattattttgaaaatattcattcgatacctgggtattggtaagctaaagcacagctatg
 10 attattggggagattgggtgatattaagcgatttaaatacaataaacaactcaatgctttt
 gttggcattgatatacaacgatatacaatcagggtacacacactgtagagataccatcaac
 aagcgtggtaataaaaaagcgagaaaactttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgctgattattactacaaactaagaaagcagcct
 aatgagaaacctcataagactgccatcattgctgtataaatcgattattaaaaacaatt
 15 cattatcttgtaataatcataaattgtacgattatcaaatgtcaccacattag

Sequence 1220

MERFCCVNQINYIQMNPLEAKFKTSALRSWKTQADAHKLACLGPTLKQTDNLPHELIF
 FELRERVRFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSIIALNIAEIFTHPDMV
 20 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLRLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFSKNQQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKKARKLLEWVIMNIIRGQHHYDNHVVDYKYLRKQP
 NEKPHKTAIIACINRLKTIHYLVMNHKLYDYQMSPH*

25 Sequence 1221

Contig_0572_pos_1540_1070,
 is similar to (with p-value 3.0e-36)
 >sp:sp|P37568|YACG_BACSU HYPOTHETICAL 17.7 KD PROTEIN IN LYS
 S-MECB INTERGENIC REGION. >gp:gp|D26185|BAC180K_145 B. subti
 30 lis DNA, 180 kilobase region of replication origin. NID: g46
 7326. >gp:gp|Z99104|BSUB0001_83 Bacillus subtilis complete g
 enome (section 1 of 21): from 1 to 213080. NID: g2632267.
 gtgatatctatgcacaatatgtccgacatcatagaacaatacattaagcgggtatttgaa
 gaagcagatgaagatggttagaaatacaacgcgctcatattgctcaacgtttcgattgt
 35 gttccttctcaacttaactatgttattaagacacgttttactaatgaacatggttatgaa
 atagaaagtaaacgtgggtggcgtggttacattcgaatcactaaaattgaaaaataaagat
 gctacaggttatattaatcacttactacaattaataggtccatctatttctcaacaacaa
 gggattattgcatagatggtttgtagataaagggttgatcaatgaaagagaagctaaa
 atgatacagacattatttgatagagaaactttaaaaatggatgttggtgcacgcgatatt
 40 attagagctaataatcttaaaaacgattactaccagttattaattattactag

Sequence 1222

VISMHNMSDIEQYIKRLFEEADEDVVEIQRAHIAQRFDCVPSQLNYVIKTRFTNEHGYE
 IESKRGGGGYIRITKIENKDATGYINHLLQLIGPSISQQQGYVIDGLLDKGLINEREA
 45 MIQTIIDRETLKMDVVARDIIRANILKRLLPVINY*

Sequence 1223

Contig_0572_pos_1064_486,
 is similar to (with p-value 7.0e-19)
 50 >gp:gp|U40604|LMU40604_2 Listeria monocytogenes ClpC ATPase
 (mec) gene, complete cds. NID: g1314293.
 gtgaggtgtttaaattgctttgtgaaaattgccattttaatgaagcggaagttaaactt
 actgttaaaggatagatagatagatgacgatgaaaaatgggtatgttcagtatgtgccaaagga
 gaaaacccctgggttacattctaacgatgataatacgtatcatacacaccaagacgatata
 55 gaagaagcattttagtgtaaacagatacttcaacaccttgctgcaaaacatggtattaat
 tttcatgagatggcatttaagaagaaaaaaatgcccaacgtgtcagatgacacttaag
 gatattgcacatgttggtgaagcttgggtgtgctgattgttatgctacgtttaaagaagac
 atcattgatatagttcaacgtgttcaaggtggtcaatttgaacatgtaggaaaaaacacca
 caatcatcgatataagaacttgcaataaaaaagcaattgaagaaaaatcaaaatatcta

aataaattgatagatgggtcaagagtttgaagaggcagcgattgttcgtgatgaaattaa
gctttaaaaagtgaagagcgaggtgtctcatgatgagtaa

Sequence 1224

5 VRCLKLLCENCHFNEAEVKLTVKGIDSTHEKWVCSVCAQGENPWLHSNDDNTYHTHQDDI
EEAFVVKQILQHLAAKHGINFHEMAFKEEKKCPTCQMTLKDIAHVGLGCADCYATFKED
IIDIVQRVQGGQFEHVKGTPQSSYKKLAIKKQIEEKSXYLNKLIDGQEFEEAAIVRDEIK
ALKSESEVSHDE*

10 Sequence 1225

Contig_0572_pos_0_454,

is similar to (with p-value 2.0e-27)

>sp:sp|P37570|YACI_BACSU HYPOTHETICAL 41.1 KD PROTEIN IN LYS
S-MECB INTERGENIC REGION (ORFX). >gp:gp|D26185|BAC180K_147 B
15 . subtilis DNA, 180 kilobase region of replication origin. N
ID: g467326. >gp:gp|Z99104|BSUB0001_85 Bacillus subtilis com
plete genome (section 1 of 21): from 1 to 213080. NID: g2632
267.

atgtctgaggagacacctgttattatttcttccagaattcgattagctagaaatcttgaa
20 aacctgtccaccacttatgttcccttcagagcaagaaggatatcgagtataaatgaa
gttcaagatgcgcttttcaacttaactttaaatcgattagatcgatggatcaacaaagt
aaaatgaaattggttgcaaacatcttgtgagtcctgaactagtgaacaacctgcttca
gcagtaattgttaaatgatgatgaatcggttaagtgttatgataaacgaagaagatcatata
cgaatacaggctctaggaactgatttatcgctaaaggatttatcaacgcgcttctaaa
25 attgatgatgaattagataaaagcggttagacattagttatgatgagcatttaggatattta
actacctgtcctactaataattggtacaggaatgc

Sequence 1226

MSEETPVIISSRI LARNLENHVHPLMFPSQEGYRVINEVQDALSNTLNRLDTMDQQS
30 KMKLVAKHLVSPPELVKQPASAVMLNDDSVSMINEEDHIRIQALGTDLSLKDLYQRASK
IDDELDKALDISYDEHLGYLTTCPTNIGTGMX

Sequence 1227

Contig_0573_pos_9024_8332,

35 is similar to (with p-value 2.0e-91)

>sp:sp|Q53726|PCRB_STAAU PCRB PROTEIN. >pir:pir|S39922|S3992
2 pcrB protein - Staphylococcus aureus >gp:gp|M63176|STAPCRA
_1 Staphylococcus aureus helicase required for T181 replicat
ion (pcrA) gene, complete cds. NID: g153060.

40 atgtacgacataaccaagtggaaacatatgtttaaattagatccggctaaatcaatttcg
gatgaaaatttagaggcactgtgtatgtctaacactgatgcaataattattggtgggaca
gatgatgtaacagaagataatgttattcatttaattagtagagtaagacggttatccgtta
ccacttctcttagaagtttcgaatgtagaaagtgtgatgcctgggttttgatttctatttt
attccaacagtcataatagtaaggatacaaaatatacataacgaaatcttactagaagcc
45 cttaaaaaatatggacatgtgattaattttgatgaagttttttcgagggatatgtcggt
ctaaacgcaaatagtaaaagttgcaaaaattaccaagcttatactcaattaggtatagaa
gatgtcgaagcatatgcacaaatggcagaagaattatatacgatttccaatcatgtacgta
gaatatagtgccacatatggagatgttgataaggttaaagcgattgcaaatatgcttcaa
catactcaattattttatggcgggtgtgataacaaacattgacaaagctaacgaaatgtct
50 aacattgcggtataccattgttgcggcgatattatatataacgacattaaaaaagcat
aaaactgtaaagataaaggaggtctaataaatga

Sequence 1228

55 MYDITKWKHMKFLDPAKSISDENLEALCMSNTDAIIIGGTDDVTEDNVIHLSRVRRYPL
PLVLEVS NVESVMPGFDYFIPTVMNSKDTKYHNEILLEALKKYGHVINDFEVFFEGYVV
LNANSKVAKITKAYTQLGIEDVEAYAQAEELYRFPIMYVEYSGTYGDVDKVKAIANMLQ
HTQLFYGGGITNIDKANEMSNIADTIVVGDIYNDIKKALKTKVIKESNK*

Sequence 1229

Contig_0573_pos_8314_6146,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q53727|PCRA_STAAU ATP-DEPENDENT HELICASE PCRA (EC 3.6
 .1.-). >pir:pir|S39923|S39923 DNA helicase pcrA - Staphyloco
 ccus aureus >gp:gp|M63176|STAPCRA_2 Staphylococcus aureus he
 licase required for T181 replication (pcrA) gene, complete c
 ds. NID: g153060.

5 atgaattcagagcaaaagtgaagcgggttagaacaacagaaggccattgcttattatggca
 ggtgctggatcaggaaagacacgtgtgttaacacatcgcatcggtatttattagatgaa
 aaagatgtatcaccttataatatttttagctattacgtttacaaataaagcagctaaagaa
 10 atgaaggcgcgtgtcgaacatcttgtgggagaagaagcgcaagtgatttggatgtccact
 tttcactctatgtgtgtaagaattctgagaagagatgctgatcggtattggcattgaaaga
 aatttcactatcattgatcctaccgatcaaaaatcagtgattaaagatgtattgaaaagt
 gaaaatatagacagtaagcgatttgagccacgtatgtttattggtgcaattagcaatttg
 15 aaaaatgaattaaaaacacctgaggatgctcaaaaagaggcgcaatgattttcactctcaa
 atgggtgcaacgggtttacaaagggttatcaaaagacagttatcacgtaatgaagcactcgac
 tttgatgtattatgacaactattaattttatttgaacgtgtaccgaaactctagaa
 tactatcaaaaataaatttcaatatatacatgtagatgagtatcaagataccaataaagca
 caatataccttagtaaaactattagcaaaacaaatttataaaatttatgtgtgttgggtgat
 20 tctgaccaatctattttatggttggagaggagctgatatacaaaatattttatcttttgaa
 gaggactatcctgaggcaaaagacaattttcctcgaacagaactatcggttcaactaagaat
 attttaaatgctgcaaatgaagttataaaacataattctgaacgtaaacctaaaggtcta
 tggactgcaaatcttgaggagacaaaattcagttattatgaagctatgactgaaagagat
 gaagcagaatacgtttgttaagaaataatgaagcatcaacgcagtggttaaaaaatatagt
 25 gaaatggctatattatataagaacaaatgcccaatcacgtgtacttgaggaaacatttatg
 aaatcaaatattccttatacaatgggtgggggtcaaaagtcttatgaccgtaaagaaatt
 aaagatttacttagttatttaagagttattgctaataagcaatgatgatattagtttgcaa
 cgtattattaacgtgcctaaacgtggtattggacacctcatctgttgaaaaaatccaaacc
 tatgcacttcaaaaataatataagtatgtttgacgcattggctgaggtagattttataggt
 30 ctctctaaaaaggttaactcaagaatgtatcagtttttatgaaatgattcaaaattaatc
 aaagacaagaattttctcgaaattagtgaatcgtagatgaagtactacaaaatcaggc
 tatagagacatgcttgatcgagaacaaagtattgaatcacgaagtcgattagaaaactta
 gatgaatttatgtctgtacctaagattatgaggaaaataactccttttagagggaacaatca
 cttattaattttctaacagatttatcattagttgctgatattgacgaagcagatacacag
 35 aatgggtgaacattgatgacaatgcattcagcaaaaggtccttgaatttcctatagttttt
 attatgggaatggaggagtcgttgttccacatatcagagcaataaaaaagtgaagatgat
 catgaaatggaagaggaaacgtcgattttgttatgtagcaattacacgagcagaagagttg
 ctttatatcacaaatgcacgaccagaatgttgtttgggtcgttctcaatccaatatgcc
 tctcgatttttaaaagaaatcccagaagacctacttgatagtcataccgggtcaaaaaaga
 40 caaactatatctcccaaatctcaacctaaaagaggttttagtaagcgctactacatcaact
 aaaaaacaagtttcatcatctgattgaaagtaggagataaagttatgcataaagcatgg
 ggtgaagggatggttagtaacgtgaatgaaaaaaatggatctgtagagttggatattata
 tttaaatcagaaggtccaaaacgattattagctcagttcgaccaataacaaaagaaggag
 gactcatag

45 Sequence 1230
 MNSEQSEAVRTTEGPLLIMAGAGSGKTRVLTHRIAYLLDEKDVSPYNILAITFTNKAKE
 MKARVEHLVGEEAQVIWMSTFHSVCVRILRRDADRIGIERNFTIIDPTDQKSVIKDVLS
 ENIDSKRFEPRMFIGAISNLKNEKTPEDAQKEANDFHSQMVATVYKGYQRQLSRNEALD
 50 FDDLIMTINLFRVPEPTELEYQNKFEQYIHVDEYQDTNKAQYTLVKLLANKFKNLQVVG
 SDQSIYGWRGADIQNILSFEEDYPEAKTIFLEQNYRSTKNILNAANEVIKHNSERKPKGL
 WTANSGGDKIQYYEAMTERDEAEYVVKIEMKHQRSGKKYSEMAILYRTNAQSRVLEETFM
 KSNIPYTMVGGQKFYDRKEIKDLLSYLRVIANSNDDISLQRIINVPKRGIGPSSVEKIQT
 YALQNNISMFDALAEVDFIGLSKKVTQECISFYEMIQNLIKEQEFLEISEIVDEVLLQKSG
 55 YRDMLDREQSIESRSLNLFDFMSVPKDYEENTPLEEQSLINFLTDLSLVADIDEADTQ
 NGVTLMTMHSAGLEFPFVIMGMESLFPHIRAIKSEDDHMEEEERRICYVAITPAEEL
 LYITNATTRMLFGRSQSNMPSRFLKEIPEDLLDSHTGQKRQTI SPKSPQKRGFSKRTTST
 KKQVSSSDWKVGDKVMHKAWGEGMVSNNVNEKNGSVELDIIFKSEGPKRLLAQFAPITKKE
 DS*

Sequence 1231

Contig_0573_pos_6142_4145,

is similar to (with p-value 0.0e+00)

5 >gp:gp|AJ011676|BST011676_1 Bacillus stearothermophilus lig
gene. NID: g3688228.
atgcaagatggttaaaaagcgtgtggaaaaattacatgacttattgaatcaatatagttat
gaatattatgtacaagataatccctcagtcctgacagtgagtgataagttattacat
gagctgattgaaattgaagaaaaatatccagaattcaaactgacagactctccaacagtg
10 cgtgtgggtggcgaagctcagtccttcttttgaaaaagtaaatcacgacacgcctatgta
agtttaggtaatgcttttaagaagaagatttaagaaaatttgatcaacgtattcgtgat
agtattggtaaggtcgaatacatgtgtgaacttaaaatagatggtttggtgctgttcgctc
aaatatgaaaatggctcgttttgttcaaggacttacacgtggtgatggtagcagagtgag
gatatactgaaaatctaagaactatacatgctataccactaaaaattaaagaacctctc
15 aattttgaggtccgtgggaagcttatatgccacgtcgttcattcattcatttgaataat
gaaaaagaacaaaatggtgaacaaccttttgcaaatccacgaacgctgctgcaggctct
ttaagacaacttgactctaaactagctgcgaaaagaaagtttaagcgtcttcttatatagt
gtgaatgacctaacgagtttaatgcaacaacacaaagtgaagcglagaggaattggac
caattaggttttaaaactaaccaagaacgtgaacgagtgatcagatattgagggcgtactt
20 aattatagagaaatggacaagcaaaagaggtatctttatctacgatattgatggatt
gttataaaagttaacgatttatctcaacaagaggaaatgggttatacgcaaaaatctcca
agatgggcgattgcttataaaatttccagctgaagaagttattacaaaattattggatatt
gagctaagtattgggcgtacgggtgttgtagaccaactgcaattctagaacctgtaaaa
gtagctggtactacagtttcaagagcctcacttcataatgaagatttaatacatgaaaga
25 gatatacgtatcggagatagtgttggtattaaaaaagccgggacatcatccctgaagtt
gtaaaaagtattttagatagacgacctaacgaatcggaatttatcatatgccaacacat
tgtcctagtgtggacatgaattagttcgtattgaaggagaagttgctttacgttgatt
aatccaaaatgtcaggcacagcttattgaaggacttatacatttcgtttcaagacaagcg
atgaatatagatgggttaggtactaaaattattcatcagctatacgaaaatcagttaatc
30 aaagatgtcgcagatatttctatttgaagaagaagatttattaccattagagcgaatg
ggaaagaagaaagttgataatcttttattagcgatagaaaaatctaaagaacagtcatta
gagcatttattatttggacttggtatttagacatttaggtgtaaaagctagtcaagtaact
gtcagcgatattgaaacgatggatcaactttttaaagtaactgaaagtgaatttaattgaa
attcaagatatattggagataaaacttgcacaatctggtgtaacatatctcgaaaatagtgat
35 attcgttcattaattgaaaaattaagtaataaaaatgttaatatgtcttataaaaggaatt
aaaacaactgaaatcgaaaggtcatcctgatttttagtgggaaaacaattgtattacaggg
aaactcgagcaaatgacgagaaatgaagcatctgaatggttgaagaatgcaaggtgctaaa
gttacaagcagcgtgactaaaagtactgatattgtcatagctggagcagatgcagggtct
aaattagccaaagctgagaagtatggtactgaaatttggactgaagcagcatttattgaa
40 aaacaaaatggaatctaa

Sequence 1232

MQDVKKRVEKLHDLNQQSYEYYVQDNPSVPDSEYDKLLHELIEIEEKYPEFKSTDSPTV
RVGGEAQSSFEKVNHDTPMLSLGNFNEEDLRKFDQRI RDSIGKVEYMCELKIDGLAVSL
45 KYENGRFVQGLTRGDGTTGEDITENLRTIHAIPLKIKEPLNFEVRGEAYMPRRSFIHLNN
EKEQNGEQPFANPRNAAAGSLRQLDSKLAARKKLSVFLYSVNDLTFNATTQSEALEELD
QLGFKTNQERERVSDIEGVNLYIEKWT SKRGSLSYDIDGIVIKVNDLSQQEEMGYTQKSF
RWAIAYKFPAEEVITKLLDIELSIGRTGVVTPTAILEPVKVAGTTVSRASLHNEDLIHER
DIRIGDSVVIKKAGDIIPEVVKSILDRPNESEIYHMPHCPSCGHELVRIEGEVALRCI
50 NPKCQQLIEGLIHVSRQAMNIDGLGTKI IHQLYENQLIKDVADIFYLKEEDLLPLERM
GKKKVONLLLAIEKSQESLEHLLFGLGIRHLGVKASQVLAERYETMDQLFKVTESELIE
IQDIGDKLAQSVVTYLENSDIRSLIEKLSNKNVNMSYKGIKTTEIEGHPDFSGKTIVLTG
KLEQMTRNEASEWLKMQAKVTSSVTKSTDIVIAGADAGSKLAKAEKYGTIEWTEAAFIE
KQNGI*

55

Sequence 1233

Contig_0573_pos_3735_3064,

putative peptide of unknown function

atgagcgaaaaagaaaaagaaaaagcaaaaatgctaatagagaatcttggactcaatccatat

tctacagataaggggaatgtcaatggataaagctacaaaatatgcacttcaattaaggggtg
 attgctcaagaaagctatcctaattgctgatagacattcctttctagtcgaaaaattacgc
 ttacttattcaacaattaaaacaattctattcatcatctcaacaattagatgatgccatg
 attcaatttagcacaacaactcgattattttggaaaatattcattcgataacctgggtattggt
 5 aagctaagcacagctatgattattggggagattgggtgatattaagcgattttaaatacaaat
 aaacaactcaatgcttttggcattgatatacaacgatataatcaggtcatcacacac
 tgtagagataccatcaacaagcgtggtaataaaaaagcgagaaaacttttattttgggtg
 attatgaatataataagagggcagcatcattatgacaatcatgctgctgattattactac
 aaactaagaaagcagcctaataagaaaacctcataagactgccatcattgcttgataaat
 10 cgattattaaaaacaattcattatcttgtaataatcataaattgtacgattatcaaatg
 tcaccacattag

Sequence 1234

MSEKEKSKSNANENLGLNPYSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLR
 15 LLIQQLKQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLTAMIIGEIGDIKRFKSN
 KQLNAFVGIDIKRYQSGHTHCRDTINKRGNKKARKLLFWVIMNIIRGQHHYDNHVVDYYY
 KLRKQPNKPKHTAIACINRLKTIHYLVMNHKLYDYQMSPH*

Sequence 1235

20 Contig_0573_pos_2231_1056,
 is similar to (with p-value 2.0e-41)
 >gp:U09991|SVU09991_1 Streptomyces venezuelae ISP5230 chl
 oramphenicol resistance protein (cmlv) and chloramphenicol
 phosphotransferase genes, complete cds. NID: g498886.
 25 atgaaagataataaaatgtgttcattattttatgataggaacatttacagtaggaatg
 gctgaatatgtagtgacaggattacttacacaaatcgctgacgatataaggtttctatt
 tcgagtgacaggtttattaattagtgtttatgctattagtgttcattgataggcccttta
 atgcgaatcataacattgaaagttcacgcccaccgtctgtaccgatttttagttgcgatt
 tttataataagtaatttagtgggaatgttagcaccgaattttaatgtattgttattatca
 30 agactcatgtctgcggcaatgcatgcgccattcttcgggtgtgtgatagtggtgtgcg
 acagtcgcacctctgtctaaaaaaacacaggccattgcacttggtcaggcaggtttaact
 attgctgtaattgttaggtgtaccattcggaatcatttttaggtggccttgcaattggaga
 gttgtttttggattcatgattgtgttggaatcattactatgttaggaatgattaaattt
 gttccaaatgtttctttaagtcagaagcaaatattagcaagaattaacagtgtttaag
 35 aatccacacattttaattgtgattgcaattattgtgtttggttactctggtgtgtttact
 acttatacatttatggagccaatgatacagagatttttctccatttaaaattgtagggtta
 actgtttgtttattatgtttggtctagcggtgtgatagggaatttaattactggtaat
 gtaccggaagataaattaacaaaaaatttataccttacatttcttttactatttgtaaca
 atcatactatttggttactgttattcaaaaatcaatattagcattaatcatttgcttctta
 40 ttcgggttttggtacatttggtacaacaccgttacttaatagcaaaattatcttaagtga
 aaagaagcaccacttctgcaagtacgttagctgcttctattttcaatgttgctaatttt
 cttggtgcaatcattggatctatattatcaatagggttaccttacattcaaaattact
 ttgatatctggtgggattatagtgttgggtatgcttcttaattcttgtaatacaactttat
 gaaaaaacaatatcacatttaataatgaattcatga
 45

Sequence 1236

MKDNKMLFIIIFMIGTFTVGMAEYVVTGLLTQIADDMKVSISSAGLLISVYAISVALIGPL
 MRIITLKVHAHRLPILVAIFIISNLVGMLAPNFNVLLSRLMSAAMHAPFFGVCMSVAA
 TVAPPAKKTQAIALVQAGLTIIVMLGVPEGSFLGGFANWRVVFVGMIVLAIITMLGMIKF
 50 VPNVLSAEANISKELTVFKNPHILIVIAIIVFGYSGVFYTFMEPMIRDFSPFKIVGL
 TVCLFMFGLGGVIGNLITGNVPEDKLTKNLYLTFLLLFVTIILFVTVIQNSILALIICFL
 FGFGTFTGTPLLNSKIILSAKEAPLLASTLAASIFNVANFLGAIIGSILLSIGLPYIQIT
 LISGGIIVLGMLLNLVNLQLEYEKKHITFNEYS*

55 Sequence 1237

Contig_0575_pos_141_485,
 putative peptide of unknown function
 gtgacaaaccggaggaaggtggggatgacgtcaaatacatcatgcccttatgatttgggc
 tacacacgtgctacaatggacaatacaaaagggcagcgaaaccgcgaggtcaagcaaatcc

cataaagttgttctcagttcggattgtagtctgcaactcgactatatgaagctggaatcg
ctagtaatcgtagatcagcatgctacggtgaatacgttcccgggtcttgtacacaccgcc
cgtcacaccacgagagtttgtaacacccgaagccggtggagtaaccatttggagctagcc
gtcgaaggtgggacaaatgattggggtgaagtcgtaacaaggtag

5

Sequence 1238

VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSSDCSLQLDYMKLES
LVIVDQHATVNTFPGLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVVTR*

10

Sequence 1239

Contig_0575_pos_4020_0,

is similar to (with p-value 1.0e-63)

>sp:sp|P00497|PUR1_BACSU AMIDOPHOSPHORIBOSYLTRANSFERASE PREC
ursor (EC 2.4.2.14) (GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE A
MIDOTRANSFERASE) (ATASE). >pir:pir|A00582|XQBS amidophosphor
ibosyltransferase (EC 2.4.2.14) - Bacillus subtilis >gp:gp|J
02732|BACPURF_7 B.subtilis pur operon encoding purine biosyn
thesis enzymes, 12 genes. NID: gl43363. >gp:gp|Z99107|BSUB00
04_97 Bacillus subtilis complete genome (section 4 of 21): f
rom 600701 to 813890. NID: g2632866.

15

20

atggagaacggtgcttatatttttagcaagtgaacatgtgcgattgatgttttaggtgct
gaatttatacaagatattcatgcaggtgagtatgttgttattacggatgaaggtagaa
gttaagacttacacacgacaaacaactgcaatttcagctatggaatatattttttt
gcgagacctgattcaacgattgcaggaaaaaatgttcatgcggtacgaaaggcatcaggt
aaacggttagcacaggaaaaaccagcaaaagcagatatggtaataggcgtacctaattca
tcattatctgcagcaagtggttatgctgaagaaataggcctaccatatgaaatgggacta
gttaaaaatcaatatgttgctcgaacttttatacaacctactcaggaattaagagagcaa
ggtgtacgtgtgaaactgtcggctgttaaggatatgttgatggttaaagatatcgtactt
gtagatgattcgattgttcgaggtacaacgattaacgcatagttaaaatgcttaaggat
tcaggagctaaccgcattcacgtaagaattgcttctccc

30

Sequence 1240

MENGAYILASETCAIDVLGAEFIQDIHAGEYVVITDEGIEVKTYTRQTTAISAMEYIYF
ARPDSTIAGKNVHAVRKASGKRLAQENPAKADMVIGVPNSSLSAASGYAEEIGLPYEMGL
VKNQYVARTFIQPTQELREQGVRVKLSAVKDIVDGKDVLVDDSIVRGTTIKRIVKMLKD
SGANRIHVRIASP

35

Sequence 1241

Contig_0577_pos_8441_9043,

putative peptide of unknown function

atgaaaaatgtttctaaagctttgatttggtttgttataagcttcatcatctttcacgca
atattatttgtgatgtggggagacatcaagaatactggtatttatatactggcattatg
ttaatagctggaataagttatgttttttaccaaagagacattgcattctaaacgattatta
acttccataggcatgggtataataacgagtgctgcacttattattatacaattaattttt
tcacttatttcatcagaattatcatacgcattctttaaatacaagaattatcacgaacgggt
gtctactttaaatggcaaatgctcgttactttattatttgtgataccttgcattgaatta
tatatgagaactgttttcaaaaagggaattaataaaaataaacttaccgaaatgggctagc
attttaattgttgcaatatgttcaagttcattatttatatacttagataaattgggtgatt
gtattctttattttttagctcaattcattctatctcttagctatgaatatacagagacgt
attgctacgactacaattgggtcaaatgtggctatcattttattattgatattccacgga
taa

50

Sequence 1242

MKNVSKALIWFVISFIIFHAILFVMWGEHQEYWYLYTGIMLIAGISYVFYQORDIASKRLL
TSIGMGIITSVALLIIQLIFSLISSELSYASLIKELSRGTVYFKWQMLVTLFVIPCHEL
YMRTVLQKELIKYNLPKWASILIVAICSSSLFIYLDNWWIVFFIFVAQFILSLSYEYTRR
IATTTIGQIVAIILLILFHG*

55

Sequence 1243

Contig_0577_pos_9746_10939,
 is similar to (with p-value 2.0e-65)
 >sp:sp|P37487|YYBQ_BACSU HYPOTHETICAL 34.0 KD PROTEIN IN COT
 F-TETB INTERGENIC REGION. >gp:gp|D26185|BAC180K_14 B. subtil
 5 is DNA, 180 kilobase region of replication origin. NID: g467
 326. >gp:gp|Z99124|BSUB0021_160 Bacillus subtilis complete g
 enome (section 21 of 21): from 3999281 to 4214814. NID: g263
 6442.
 atggaccatagttccgcttcgaaaaaattaattaagatatagagcaaaatcagtatgta
 10 acagtaaaacatttatctcatgatgatttttatattgatgatttggtaaaaaagaaagaa
 gtcattgcaagccttgaaatacctaaggattttctcaaaacaccttaagataatgattta
 aataagactcttcattatatagcagagatgattttataggacatattgctatggaaata
 atcagtcgatcattatacgaacaacaaatccctaattatttcatgaacatcttgatgat
 atgaaacaaccacaatccttagataaaagtgaacaatcttattattcgcttacacctcaa
 15 tctaaaaataaaaagtgtagctatcaataaacatgctcatcaatccatttcaattggcatt
 gtatttgggtgcgtcatctttgtaagtgttatccaaatcctattacatcaacgtcttaaa
 cgaacgcacctctcgaagattatatttggtagcttatagtcacttaaaactatacttg
 acttatatcagtgtagacagaattttctaactttgaaacagctggccctttatactataga
 gcagaaccagttgggtgtagtgcaacaattttatataaaaatgtataaagaacgtggattt
 20 gaaattaaaccagaaatcgctggacttatgatctcagctataatttctgatagtttatta
 tttaaatcacctacctgcacaaaagaagatgtagatgctgctcaagcactaaagatatt
 gcaaattgttgatttagaagcatatgggttagaaatgttaaaagcaggtgcttcaactaca
 gataaatctgctgaaacacttgctcaatatggatgctaaatcattcaatatgggagattat
 gtaacacgtattgtctcaagtcaatactgtagatattgatgaagtttttagatcgtaaagag
 25 gaatttgaaaaagttatgttagaaatgagtgccaatgaaaaatacagatttattcgcttctt
 gtagttactgatattattaaacagcgattctaaaatccttgtagttgggtgctgaaaaagat
 aaagttggagaagcattttaagtagaactagatgatggatggctttcttatctggcggt
 gtatcacgaaaaaaacaagttgttcctcaaatcactgaagttttaactcaataa
 30 Sequence 1244
 MDHSSASKKLIKDIEQNQYVTVKHLSHDDFYIDDLVKKKEVIASLEIPKDFS KHLKDN DL
 NKTLP LYSRDDFI GHIAMEIISRSLYEQQIPNIIHEHLDDMKQPQSLDKVKQSYSLTPQ
 SKIKSVAINKHAHQSI SIGIVFVVVIFVSVIQILLHQRLKQNA PLERLYLVPYSQLKLYL
 TYISVHRISNFETAGPLYRAEPVGC SATILYKMYKERGF EIKPEIAGLMISAIISDSL
 35 FKSPTCTKEDVDAAQALKDIANVDLEAYGLEMLKAGASTTDKSAETLVNM DAKSFNMGDY
 VTRIAQVNTVDIDEVLD RKEEF EKVMLEMSANEKYDLFVLVVTDIINSDSKILVVGA EK D
 KVGEAFKVQLDDGMAFLSGVVSRKKQVVPQITEVLTQ*

Sequence 1245
 40 Contig_0577_pos_11239_12618,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P39616|DHA2_BACSU PROBABLE ALDEHYDE DEHYDROGENASE YWD
 H (EC 1.2.1.3). >pir:pir|S39713|S39713 hypothetical protein
 - Bacillus subtilis >gp:gp|X73124|BSGENR_59 B. subtilis genom
 45 ic region (325 to 333). NID: g413923. >gp:gp|Z99123|BSUB0020
 _91 Bacillus subtilis complete genome (section 20 of 21): fr
 om 3798401 to 4010550. NID: g2636240.
 atgacaataattagagataaaatttaacaatagtaaagctttttttaatacgcataaaaca
 aaaaaccttaaaatttcgaaaaacaacacttaaaattactaaagtaaaaaatatacaaaatcat
 50 gaaaatgaattatttagatgccttatataaaagatttaggtaaaaagtaagggtgaagcatatc
 gcaactgaaattgggtatgcttttgaaaagcataaaagctaatgcgcaaaagagttaaaaaat
 tgggtcgaaaacaaacaaacggatacaccaactctacttattccctacaaaagagttatatt
 aaaaaagaaccttacggtacggtgcttattataggaccatttaattatccggttcaatta
 gttttcgagcctctcatcgagcaatagctgccggaataactgctatagttaaaccttca
 55 gagttaacacctcatgttgccattgtgatcaaggacatcattgaagatacatttgatgaa
 gcatacgtttctgttgtgaagggtgattgaagaaacccaaacgttattaagtctacca
 ttgattatatgttctttactggcagtgaaaaagtcgaaaaattgtctatgaagctgca
 gcaagaaaatttaattcagttactctgaacttggcggtaaatcacctgtcattgtcgat
 gatacagccaatatcaaagtagcagtgtaacgtatttagtttggtaaaatttactaatgct

ggtcaaacatgtgtcgctccagattatatattagttcagcggaaagttaaaaaatgattta
 ataaaagctcttaaaaaaacaattactgaattttacggagaaaatattgaaaaaagccct
 gatttcggacggattgttaatacaaaaacacatttaacggttgaatgacttgattczaatt
 cataaagataatgtgtttttggaggttaatagttctaaagaagatttatattgaacct
 5 actttattggataacataaccaatgacaataaaatcatgaaagaagaatattcgggtccc
 attttgcctattattacttatgataatttcgatgaagtacttgaaatcatccaaagtaaa
 tcaaaaccactaagtttgatcttttagcgaagatgaaaacatgacacatagagtgggt
 gaagaattatcatttggggcggtgcaattaacgatacgttaatgcatttagctaaccct
 aacttaccttccggtggtgttaggttcttcaggcataggtcaatatcatggtaagtattct
 10 tttgatacatttagtcatatgaaatcatacacatttaaatctacacgtctagaatcgagt
 ttattttccctccatataaagggttaaatatattaaaaccttcttcaagaactag

Sequence 1246

15 MTIIRDKFNNSKAFNTHKTKNLKFRKQQLKLLSKNIKNHENELLDALYKDLGKSKVEAY
 ATEIGMLLSIKLMRKELKNWSKTKQDTPLYLFPKSYIKKEPYGTVLIIGPFNYPVQL
 VFEPLIGAIAGNTAIVKPSLTPHVAIVIKDI IEDTFDEAYVSVVEGGIEETQTLLSLP
 FDYMFFTGSEKVGKIVYEAARKLIPVTLELGGKSPVIVDDTANIKVASERISFGKFTNA
 GQTCVAPDYIILVQRKVKNDLIKALKKTITEFYGENIEKSPDFGRIVNQKHFNRLNDLIQI
 20 HKDNNVFGGNSKEDLYIEPTLLDNITNDNKIMKEIFGPILPIITYDNFDEVLEIIQSK
 SKPLSLYLFSEDENMTHRVVEELSFGGGAINDTLMHLANPNLPFGGVGSSGIGQYHGKYS
 FDTFSHMKSIFYKSTRLESSLFFPPYKKGKFKYIKTFEKN*

Sequence 1247

25 Contig_0577_pos_7994_7059,
 is similar to (with p-value 8.0e-44)
 >gp:gp|U62057|MCU62057_2 Mycoplasma capricolum NADH oxidase
 (naox) gene, partial cds, and lipoate-protein ligase (lpla),
 pyruvate dehydrogenase EI alpha subunit (odpa), pyruvate de
 30 hydrogenase EI beta subunit (odpb), pyruvate dehydrogenase E
 II (odp2), dihydroliipoamide dehydrogenase (dldh), phosphotri
 acetylase (pta) and acetate kinase (ack) genes, complete c
 ds. NID: gl480703.
 atggaagagtgatttcttaaaaatttaccttctgaagaaagttatttttatattatatt
 35 aacagaccttcaattattgttggaaagaatcagaatacaattgaagaagttatcaagcg
 tatattgataaacatcaaataagatgtagtgagacgtatttctggtggtgggctgtttat
 catgatactggaaacttaatttttagctttatcacagatgatgatggccatagctttcat
 aatttttaaaaagttttacgatgccattgtacaggccttacaatcaatgggagttaatgct
 gaaatgactggaaggaatgatatacaagtagggcaagctaaaatatctggaaatgctatg
 40 gttaaagtaaaaaatagaatgttttagtcatggtacattaatgctgaattgtgatttaa
 gaagttcaaaaaggcattaaaagtgaatccagctaaaattaaatctaaaggcgtaaatct
 gtagaaaaagagttgccaatattgaggaatttctagaacagccaatagatatagaagaa
 ttcaaaaaaattattcttaaaaactatttttggtgaaaatgaagttgaagaatatatta
 acagaagaagattggaaaaatattaagcaattaagtgtgaaaagtaacgtacgtgggaa
 45 tggaaactatggcagcaatccaaaatataatattgagcgtgaagagaaaatttgaaaaagg
 ttatttcaataaaaattagatgtaaaaaaagggaagaattgaacgggcaaaaactatttga
 gatttcttcggcgaaggagatgtaacggaactgaacatgcgttagtaggttgcttacat
 gattttgaacatatagaagaggcacttcaaaattatgatttctatcactacttggggat
 atagataagtatgaaattataagattgatgtcctaa

Sequence 1248

MEEYVLKNLPSEESYFLFYINRPSIIVGKNQNTIEEVNQAYIDKHQIDVVRISGGGAVY
 HDTGNLNFESFITDDDGHSFHNEKFTMPIVQALQSMGVNAEMTGRNDIQVGQAKISGNAM
 VKVKNRMFSGHTLMLNCDLNEVQKALKVNPAAIKSKGVKSVRKRVANIEEFLEQPIDIEE
 55 FKKIILKTIIFGENEVEEYIILTEEDWKNIKQLSDEKYRTWEWNYGSNPKYNIEREKEKFEKG
 FIQIKLDVKKGRIERAKLFGDFFGEGDVTELEHALVGCLHDFEHIEEALQNYDFYHYFGD
 IDKYEIIRLMS*

Sequence 1249

Contig_0577_pos_6503_5973,

putative peptide of unknown function

atggtaatacgcactaaatatgatgaatatcagcaaaacaatgggtactgaaattatcaga
tttgatcaaaactcgcaaagaaagtcattttaaagtacagagaattatcgaaagatcatgt
5 aaattttatggtaataattatattagtaaaaaagcagaaacgaatcgtattactggaatc
tctagtaaaaccacctattttacttacgcctcttttctacttactttttccaactcac
tcagaccgtcaagaagaaaaatatatggattaatatgcattatattgaaaatgttaaaagaa
cttaaaaatcgtaagagtaaaaaaatttttgcgaatgggtgattcggttaacgctcaatgta
tcatttcatagcttgtggcatcaatatacgaatgcaatcatctattattacatggtagat
10 aagcaatcacgaatgaaatctaacaaccctgaacaaccattgactataatcagtcctct
ctaaatattttcgaggcgctctcacgctactccctttttgaagaaaattag

Sequence 1250

MVIRPKYDEYQQTNGTEIIRFDQTRKESPFKVQRIIERSCKFYGNNYISKKAETNRI*GI
15 SSKPPIILLTPLFPTYFFPTHSDRQEENIWINMHYIENVKELKNRKSIIIFANGDSLTLNV
SFHSLWHQYTNAILIYYMVDPKQSRMKSNNPEQPIDYNQSSLNIFEALSRYSLFEEN*

Sequence 1251

Contig_0577_pos_3724_3383,

20 putative peptide of unknown function

gtgattgcggtattattttgattggtttgtatttttagtagtattttaatttctgtaact
tttaataacattttgatctatctttagctattatattaattggtagtagaatgagagca
tttgataatttaaatgcatgaagcttgcacgcattttatttacaataagttttggaat
aaatggataaacttgtttatttggcatttccagttattactagttatacagcatatcga
25 aatgctcatcattgtactggattttttaaaactacattgcctggcttttatagtgtagt
aaagataattttacaccgaactatgaaaggaaagagagtttag

Sequence 1252

VIALLFDFWVIFSSILISVTFNNILIYLVAILIGSRMRAFDNLMHEACHRSFLTNNKFWN
30 KWITCLFVAFPVFTSYTAYRNAHHCTGFKTTLPGFYSVVKDIFTPNYERKES*

Sequence 1253

Contig_0577_pos_3179_2763,

putative peptide of unknown function

atgccaaactttacaactttttattttacatcatcgggtttttgtttttgatgtatatggca
35 tggctctttatggacggaaaagccttcaaatatcgaagaaatagaacctatgtcagctaaa
aagcagattttatttgcctttatctgtttctttatttaaatcttcagcaataatggatact
gttggtgttatcggaacgagtgcttcagtttatgatggttatgacaaagttgtttttca
ttggctacaattttctgtatcttggatttgggttgccttttttagctatttttaggaagaatt
40 acaggaaaaattgataaaagtggtaagtatatcggtatttttaataaagtttctagcgtt
attgttattattgtaggattaattatattaaaaaatattgttggaattttaagtttag

Sequence 1254

MPTLQLFIYIIIGFLFLMYMAWSLWTEKPSNIEEIEPMSAKKQILFALSVSLLNLHAIMDT
45 VGVIGTSASVYDGYDKVVFSLATISVSWIWFVFLAILGRITGKIDKSGKYIVILNKVSSV
IIVIIVGLIILKNIVGILS*

Sequence 1255

Contig_0577_pos_1303_881,

50 putative peptide of unknown function

gtgaatgtgagatatcggtcggttttatgagcatatttgggtatatgagaaagacgggggt
gttgctggatgtgtaattgcgtatcctgggaaagaggaaatggattttgaacaacaatgg
cttaagttaccacttgaagaagatatccttcagttaggtacaccattacctgaaaaagaa
tcatacagatgatgaaatatatagaagcagtagtaacgactccaaaatatcgaggacaa
55 ggtattgcaacacaacttttaaaagtatgtaattccactcatgcacatgaaaaatgggga
ttgaattgtgattatgataataataaagcacgccacctatatcacaattagggtttaaa
gaggatgcgacaattcggtttatatggccatcaatattttcatatgacattgaataataag
taa

Sequence 1256

VNVRYRSFYEHIVVYEKDGVVAGCVIAYPGKEEMDFEQWLKLPLEEDILQLGTPLPEKE
 SYDDEIYIEAVVTTPKYRGQGIATQLLKYVISTHAHEKWGLNCDYDNNKARHLYHKLGFK
 EDATIRLYGHQYFHMTLNNK*

5

Sequence 1257

Contig_0578_pos_404_2398,

is similar to (with p-value 0.0e+00)

>sp:sp|P50620|RIR1_BACSU RIBONUCLEOSIDE-DIPHOSPHATE REDUCTAS
 10 E ALPHA CHAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE). >gp:
 gp|Z68500|BSNRDYMA_2 B.subtilis cw1C, nrdE, nrdF, ymaA and y
 maB genes. NID: g1154630. >gp:gp|Z99113|BSUB0010_32 Bacillus
 subtilis complete genome (section 10 of 21): from 1781201 t
 o 2014980. NID: g2634090.

15 gtgtatcctgaagaaatccatgataaaatgatttcttttgatgatgaaatcgaaagactt
 cattatcctgttgataataatttttattttaatgttttcgaaaaatatagcgaagcagaa
 ttaattgaaattactgaatacgcacaaatcaattcacttccaatttgctagtatatgtcg
 gcaagtaagttttataaggattacgcttttaaaacgaatgataaaactaagtttcttgag
 gattataatcagcagctcgcaatcggtgcactttatttagcgaatggtaatgttaacaa
 20 gcaaaacaattttatttctgcaatgggtggaacaacgttatcaaccggcgacaccaacattc
 ttgaatgcaggtagagcaagacgtggagaacttgtttcatgtttcttattagaagtagac
 gatagcttaaatcttatcaatttcattgattcaactgcaaaacaacttagtaaaattggt
 ggtggtgtgcccattaattttatcgaaacttcgtgcacgtggagaagcaattaaaggaatt
 aaaggtgtagctaaaggtgttttacctgttgctaaggcacttgaaggtggatttagttat
 25 gcagatcaattaggacaacgtcctggcgctggggcagtgacttaaatattttccattac
 gatgttgaaagagtttttagataactaagaaagtgaatgcagatgaagatttacgtttatct
 actatttcgactggtcttattgtaccttctaaattctttgatttggtctaaagagggtaaa
 gatttctatatgtttgcacctcatacagtttaacaagaatacgggtgtgacttttagatgat
 attgatttagaaaagattacgacgatattggttgcaaacctaatatcgataaaaagaaa
 30 aaagacgcacgtgaaatgttaaatatgattgcacaaacacaattacagttctggttatcca
 tatcttatgtttaagataatgcaaacaaagtacatgcgaattcaaatattgggcaaatt
 aaaaatgagtaatttatgtactgaaatttttccaattacaagagacatcagtaattaacgac
 tatggaattgaagatgaaattaaaagagatatttcatgtaacttaggctcattgaaata
 gtgaatgttatggaatcaggtaaattcagagattctgtgttcacaggtatggatgctctt
 35 acagttgtaagtgtatgaagcaaatattcaaaatgcaccaggtgtaaaaaaagcgaatagt
 gaactacattctgttgactaggagtaattgaacttacatggttatctagctaaaaataaa
 attggctatgaatctgaagaagcaaaagactttgctaataatattctttatgattatgaac
 tattattccatcgaaacttcaatggaaattgcaaaagagcgtggagaaaagtatcaagac
 tttgagcaatcagactatgcaaatggtaaatattttgaattctatacatctcaagaattt
 40 gaacaaaatttgaaaagggttcgccaactttttgatggtatcgatatacctacttcaaat
 gattggaagaattgcaaaaataaagtagaacaatatggactttatcacgcttatagatta
 gctattgctccgactcaaagtatttcttatgttcaaaatgcgacaagttctgttatgcct
 attgttgacaaaattgaacgtcgtagctatggaaatgcggaacattttatccaatgccaa
 tttttatcaccagaaacaatgtggtactacaaatcagcatttaataactgacaaaatgaaa
 45 cttattgacttagtagctacaattcaaaactcacgtagaccaaggtatttcaacgatactt
 tatgttaattcagaaatttcaacacgtgaactttcaagatttatgtatatgcacaccat
 aaaggccttaaatctttatactatacacgtaataaattattaagtgtggaagaatgtaca
 agttgtgcgatttaa

50

Sequence 1258

VYLEEIHDKMISFDDIEIRLHYLVDNNFYFNVFEKYSEAEIEITEYAKSIHFQFASYMS
 ASKFYKDYALKNTDKTKFLEDYNQHVIAIVALYLANGNVKQAKQFISAMVEQRYQPATPTF
 LNAGRARRGELVSCFLLEVDLSNSINFIDSTAKQLSKIGGGVAINLSKLRARGEAIKGI
 KGVAKGVLPVAKALEGGFSYADQLGQRPGAGAVYLNIFHYDVEEFLDTKKVNADEDLRLS
 55 TISTGLIVPSKFFDLAKEGKDFYMFAPHTVKQYGVTLDDIDLEKYYDDMVANPNIDKKK
 KDAREMLNMIAQTQLQSGYPYLMFKDNANKVHANSNIGQIKMSNLCTEIFQLQETSVIND
 YGIEDEIKRDISCNLGSNIVNVMESGKFRDSVFTGMDALT VVSDEANIQNAPGVKKANS
 ELHSVGLGVMNLHGHLAKNKIGYESEEAKDFANIFFMIMNYYSIERSMEIAKERGEKYQD
 FEQSDYANGKYFEFYTSQEFEPKFEKVRQLFDGIDIPTSNDWKELQNKVEQYGLYHAYRL

AIAPTQISISYVQNATSSVMPIVDQIERRTYGNAETFYPMPFLSPETMWYYKSAFNTDQMK
LIDLVAITQTHVDQGISTILYVNSEISTRELSRLYVYAHKGLKSLYYTRNKLLSVEECT
SCAI*

5 Sequence 1259

Contig_0578_pos_2550_3485,

is similar to (with p-value 2.0e-84)

>sp:sp|P50621|RIR2_BACSU RIBONUCLEOSIDE-DIPHOSPHATE REDUCTAS
E BETA CHAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE). >gp:g
10 p|Z68500|BSNRDYMA_3 B.subtilis cwIC, nrdE, nrdF, ymA and ym
aB genes. NID: g1154630. >gp:gp|Z99113|BSUB0010_33 Bacillus
subtilis complete genome (section 10 of 21): from 1781201 to
2014980. NID: g2634090.
atgactaatatgttctcggcgtcaaaatatctcaciaaatgtgggtggaaacagaattttaa
15 gtatcaaaagatatagcaagttggaaaacattaacagattctgagaaaaatactttttaa
aaagcgcttgcagggttaacagggttagatcacatcaagctgatgatggtatgccatta
atcatgcttcatactactgatttaagaaagaaagctgtttattcatttatggctatgatg
gaacaaatccatgcgaaaagttattctcatatcttcactacattattaccatctagtga
accaactatttattggatacttgggttattgaagagccacattttaaataataatcagat
20 aaaattgtagaaaattaccacaaaactttggggtaaagaagcatcgatttacgatcaatat
attgctcgtgtttctagtgtattcttagaacaatttctattctattctggcttctattat
ccattatatctcgcaggacaaggaaaaatgactacgtcagggtgaaattatacgttaagata
cttttagatgaatctatacatggagtggtcacagggttagatgcacaaagctacgtaat
gagttatctgaaagtggagaaacaaaaagctgatcaagaaatgtacaaattattaaatgaa
25 ctttatgataatgaagtttcatatacacatttatttatatgatgatattggcttctgtgaa
gatgtattaaattatgttcgatataatggaaataaagcattatcgaatttaggttttgaa
ccatatttgcgaagagagagaggtttaaccctattattgaaaatgcactagatacatctaca
aagaaccacgatttcttctctgttaaagggtgatggctatacattggctttaaattgttgag
cctctacgtgatgaagactttgtttttgataattaa

30

Sequence 1260

MTNMFWRQNISQMWVETEFKVS KDIA SWKTLTDSEKNTFKKALAGLTGLDTHQADDGMPL
IMLHTDRLRKAVYSFMAMMEQIHAKSYSHIFTTLLPSSETNYLLDTWVIEEPLKYSKD
KIVENYHKLWGKEASIDQYIARVSSVFLETF LFYSGFYYPYLAGQGKMTTSGEIIRKI
35 LLDSEIHGVFTGLDAQSLRNELSESEKQKADQEMYKLLNELYDNEVSYTHLLYDDIGLAE
DVLNRYRYNGNKALS NLGFEPYFEEREFNP IENALDSTKNHDFFSVKGDGYTLALNVE
PLRDEDFVFDN*

Sequence 1261

40 Contig_0578_pos_3739_4713,

is similar to (with p-value 0.0e+00)

>gp:gp|AJ005352|SAA005352_1 Staphylococcus aureus, Sst putat
ive iron transport operon. NID: g3724154.

atgaaatttatttttaaaaggttataccttattttattttatttagtgatttttaactattgtc
45 tctttattttataggggtgagtcagctctctctaataagatattttccatttaagtgatgaa
caaataaatattttgttttcgagtcgaattcccagaacagtttagtattctactttcgggt
agttcactagctttatcaggattaattatgcaacagatgatgcaaaataaatttgtaagc
ccaacgactgctggtactatggagtgggcaaaattaggtatttttaattgtcattgttggtc
tttctaattggtcccatttttaatacaattattatttgctgttggttctaagtattgttgga
50 acgtttttattttgtccaattaattaatcttatccgtgtaaaagatgtaattctttgttcca
cttttaggcattatgatgtgtggtattttatccagttttactacatttgtagcgttgaga
accaatgcttttacaagcatttggaactgggttaactggtaactttgcagttataacgagt
ggtcggttttgaggtgtgtatctcacaaataccattacttattttggcatttgtatttgca
aatcattttactattgcaggatggtgaaaagacttttagtcataatttaggtgtaagttat
55 gaaaaaatcattaaaatagcattattcataacagcaacgctaacagcattagttgttggt
actgttggaacattaccatttttaggtttaatcgtaccgaatatcatatctatctataga
ggcgatcatttgaaaaatgctttaccacacacgcttatgctcgggtgcaattttgtttta
attgctgatattatttggtagaataattgtttacccttacgaaattaatttggttgacg
ataggtgtattttggcacaattattttcctaattcttgctaataaggttaggaaaaattat

gcaaatgagcgctaa

Sequence 1262

5 MKFIFKGYTLFILLVILTIVSLFIGVSQLSLIDIFHLSDEQINILFSSRIPRTVSILLSG
SSLALSGLIMQMMQNKFVSPTTAGTMEWAKLGILMSLLFFPNGPILIKLLFAVVLISVG
TFLFVQLINLIRVKDVI FVPLLIGIMIGGILSSFTTFVALRTNALQSIGNWLTGNFAVITS
GRFEVLYLTPLLILAFVFANHFTIAGMGKDFSHNLGVSYEKI IKIALFITATLTALVVV
TVGTLPFLGLIVPNIIISYRGDHLKNALPHTLMLGAI FVLIADIIGRIIVPYEINIGLT
10 IGVFGTIIIFLILLMKGRKNYANER*

Sequence 1263

Contig_0578_pos_4739_5656,
is similar to (with p-value 1.0e-85)
>gp:gp|AJ005352|SAA005352_2 Staphylococcus aureus, Sst putat
15 ive iron transport operon. NID: g3724154.
atggcaatttgatggcaattttttatttattggtaggtagattttgatataatttgaa
tatcaattccaaagccgtttcaaaaaatttattcttatattatttagtaggagcatcgata
ggaacgtccgtagttattttccaatctattactacaaatagattattaactccttctatt
atgggacttgattctgtgtattttattgtcaaagctctacattttttatttttaggtgaa
20 caagcaactgtgtttacaaatatctacctaattttttaatacactttgatcgcaatgggtg
tttttttcattgctatttttcaagttatatttaaaactagggtcacttttcagtctatttc
atattactagttggtgtaatattaggaactttctccgtagtattacaagttttctccaa
cttataatgaatcctgaatcttttttagcagttcaaaatgttatgtttgctagctttgaa
gcatcaaattctaaattagtaactgtttcaggcatattattagttatattaataattgtt
25 acaataacttccgaccataatttagatgttctttgttaggtagagcacaagcaattaat
ctaggtgtttcttatgagaatatgacgcgtatgtttcttattttggttgctttactcgtc
tctatatctactgcattaataggtcctgtaacatttcttggttattaactgtcaactta
gcacacgaatttatgaagacctatgaacacaaatttattttaccggcaacaatattattc
agctggattagttttatttatagcgcaatgggtagttgaaaatctattcgaagcaacaact
30 gaattcagtttgatagtagatttagttggtggaagttacttcataatttgcttgctcaaa
aggagaaatgcaaattga

Sequence 1264

35 MAICMAIFYLLVGLDFDIFEYQFQSRFKKFILILLVGASIGTSVVIFQSITTNRLLTPSI
MGLDSVYLFVKVLPFILGEQATVVTNIYLNFLITLIAMVFFSLLLFQVIFKLGHFVSVEF
ILLVGVLGTFFRSITSFLQLIMNPESFLAVQNMVFASFASNSKLVTVSGILLVILIIV
TIILRPYLDVLLGRAQAINLGVSYENMTRMFLILVALLVSISTALIGPVTFLGLLTVNL
AHEFMKTYEHKFILPATILFSWISLFIAQWVVENLFEATTEFSLIVDLVGGSYFIYLLVK
40 RRNAN*

Sequence 1265

Contig_0578_pos_5914_6432,
is similar to (with p-value 6.0e-67)
>gp:gp|AJ005352|SAA005352_3 Staphylococcus aureus, Sst putat
45 ive iron transport operon. NID: g3724154.
atgaacataactatcgagcaactagttaaatttggtcggttttccatattcaaaaggtcga
atgaagcaagaagactatgacaaagtgaatatgctcttgacttattacagcttaatgaa
attaaacatcgtaatatataaaacactatctggagggtcagcgacaacgtgcataattgct
atgactatagcacaagatactgattatatattatttagacgagcctttaaataatttagat
50 atgaacattcagtagcagattatgcaaacgttacgagatttatgtcgtcagttaaataaa
acaatcattatcgatattacatgatatcaactttgcttcttggtattcagacgacatcatt
gcgcttaaaacaaggtgagctagttaaagctgatgataaagataatgttattcagctcgac
attttaaaaagtttattatgaaatggaagtacgtatagaggagataaaggggacaacgtatt
tgtctatattatgatgaaactacttttgactcagtttaa
55

Sequence 1266

MNITIEQLVKFRFPYSKGRMKQEDYDKVEYALDLLQLNEIKHRNIKTLSSGQRQRAYIA
MTIAQDTDYILLDEPLNNDMKHSVQIMQTLRDLRQLNKTIIIVLHDINFASCYSDDII
ALKQGLVKKADDKDNVIQSDILKSLYEMEVRIEIRGQRICLYDETTFDSV*

Sequence 1267

Contig_0578_pos_6511_7554,
is similar to (with p-value 0.0e+00)

5 >gp:gp|AJ005352|SAA005352_4 Staphylococcus aureus, Sst putative iron transport operon. NID: g3724154.
atgaaaaaacagtccttatttttattattgtctctagtttttagttttaacggctttagt
aatagttcgaataataattcaacttcgaaaaagaaaaatagtgattctaaagaaactgta
accatcaaaaatagttttgaagcaagtggttaaagaaaataatggcagtgataagaaaaaa
10 atctctaatactgtcgaagtaccaaagaatcctaaaaatgccgttgatttagattatgga
gcgcttgatgtgtgaaagaattaggtgtggctgataaagtaaaagggttacctaaagg
gaaaaataaccaatctttacctaatttttagatgaatttaaagatgataagtatattaat
actggaaatttaaagaagtgaactttgataaagttgcatcagctaaaccagatgtgatt
tttatttcaggagaacagctaatcagaaaaatttagatgaatttaaaaaagctgcacca
15 aaagctaaagttgtatatgtaggtacaagtgatgacaacttaattaagatatgaaaaaa
aatacagaaaaatttagggaaaaatctacgataaagaagataaagctaaaaaaattaataa
gattctagatagaaaaatatctgatatgaaagataaaactaaagactttaataagaaagta
atgtattttatgggttaacgaaggtgaactatcaacgtttggaccaggaggaagatttgg
ggtttagtggttatacattaggatttaaacctgcagacaaaaagggttagcaaaagcccg
20 catgggtcaaaatataaataatgaatatattaacaagcagaatccagatgttatttttagct
atggatcggtgttcagttgtaggtggttaaagcaacaacaatcaagtttaaaaaacaaa
gttataaaaaatgtaaaagcagtaaaagtaatcatatttacgaattagatccaaaacta
tggtatttctcttcaggatcttcaacgacaactatcaacaaattgatgaattaaatgaa
gtagtagagaaaagttgaaaaataa

25 Sequence 1268
MKKTVLFLLLSLVLVLTACSNSSNNNSTSKKKNDSKETVTIKNSFEASGKENNGSDKKK
ISNTVEVPKNPKNAVVLVDYALDVLKELGVADKVKGLPKGENNQSLPKFLDEFKDDKYIN
TGNLKEVNFDFKVASAKPDVIFISGRTANQKNLDEFKKAAPKAKVVYVGTSDDNLIKDMKK
30 NTENLGKIYDKEDKAKKINKDLDRKISDMKDKTKDFNKKVMYLLVNEGELSTFGPGGRFG
GLVFDTLGFKPADKKVSKSPHGQINNEYINKQNPVILAMDRGSVVGKATTNQVLKKN
VIKNVKAVKSNHIYELDPKLWYFSSGSSTTTIKQIDELNEVVEKVEK*

Sequence 1269

35 Contig_0578_pos_8973_8041,
is similar to (with p-value 7.0e-41)
>sp:sp|P18579|MURB_BACSU UDP-N-ACETYLENOLPYRUVVOYLGLUCOSAMINE
REDUCTASE (EC 1.1.1.158) (UDP-N- ACETYLMURAMATE DEHYDROGENA
SE). >pir:pir|S26500|A43727 probable division initiation reg
40 ulatory protein 1 - Bacillus subtilis >gp:gp|M31827|BACDDSA_2
Bacillus subtilis (clone lambda-BS1) cell division and spo
rulation protein (dds) gene, complete cds. NID: gl42831. >gp
:gp|Z99111|BSUB0008_195 Bacillus subtilis complete genome (s
ection 8 of 21): from 1394791 to 1603020. NID: g2633699.
45 atgttcaaaacattgaataaaaatgacatcttacgcgagtagagtcattcttcctaaa
gatattattaaagtggatgaacctctcaagcgttatatacatatacagaaacaggaggagag
gcagatttttatttatccctacaaaaaatgaagaagtccaagccatcgtaaagtttgcc
catgagaacagtataccggttaacttatttaggaaatgggtctaactatcattcgagaa
gggtgaattcgaggaatcgctcctcagcttattatctctcaatcatattgaaacctctgat
50 gatgcaattatagcaggtagtggtgcagcaattattgacgtttcaaatggtgcacgtgac
catgtattaaccggtttagaatttgcatgcggtatccctgggtcaattgggtggcgcgcta
ttcatgaatgctggtgcttatggcggaaggttaaagactgtattgcatgcatgattatgt
gtcaatgaaaaagggtgatttattaaagctcactacagctgaactggaattagactataga
aatagtggtgtacaacaaaaacatttagttgtattagaggctgctttcaccttagaacca
55 ggtaaatagatgaaattcaggccaaaatggatgatcttactgaaagacgtgaatctaaa
caaccgcttgaattcccttcttgcggaagtgtttccaaagaccaccgggtcattttgca
ggtaaacctcattcaagattctaatttacagggtatcgaatcggtggcgttgaagtttca
actaagcatgcgggattcatggttaatgtagacaacgggtacagcaactgattatgaagca
cttatcatcacgtacaaaaaatagttaaagaaaaattcgatgttgaattgaatactgag

gtacgtattataggtgatcatccacagattaa

Sequence 1270

5 MFKTLNKNLILRGLESILPKDIIKVDEPLKRYTYTETGGEADFYLSPTKNEEVQAIVKFA
HENSIPVTYLGNLSNIIIREGGIRGIVLSLLSLNHIETSDDAIAGSGAAIIDVSNVARD
HVLTLGLEFACGIPGSIGGAVFMNAGAYGGEVKDCIDYALCVNEKGDLLKLTAELELDYR
NSVVQKHLVVLAAFTLEPGKLDEIQAKMDDLTERRESKQPLEFPSCGSVFQRPFGHFA
GKLIQDSNLQGYRIGGVEVSTKHAGFMVNDNGTATDYEALIHVQKIVKEKFDVELNTE
VRIIGDHPTD*

Sequence 1271

Contig_0578_pos_4304_3819,
is similar to (with p-value 2.0e-60)
>gp:gp|AJ005352|SAA005352_1 Staphylococcus aureus, Sst putat
15 ive iron transport operon. NID: g3724154.
gtgagatacaacacctcaaaacgaccactcggtataactgcaaagttaccagttaaccag
tttccaatgctttgtaaagcattgggtctcaacgctacaaatgtagtaaaactggataaa
ataccaccaatcataatgcctaaaagtgaacaaagattacatcttttacacggataaga
ttaattaattggacaaaataaaaacggttccaacaatacttagaacaacagcaataataat
20 ttgattaaaatgggaccattaggaagaacaacaatgacattaaaatacctaattttgcc
cactccatagtagcagcagtcggtgggttacaattttatgttgcacatctgttgcata
attaatcctgataaagctagtgaaactaccgaaagtagaataactggttctgggaatt
cgactcgaaaacaaaatattttatgttcatcacttaaatggaaaatatctattagagag
agctga

Sequence 1272

VRYNTSKRPLVITAKLPVNQFPMLCKALVLNATNVVKLDKIPPIIMPKSGTKITSFTRIR
LINWNTKNVPTILRTTANNNLIKMGPLGKNNNDIKIPNFAHSIVPAVVGLTNLFICIICCI
INPDKASELPESRILTVLGIRLENKIFICSSLKWKISIRES*

Sequence 1273

Contig_0580_pos_1602_1976,
is similar to (with p-value 6.0e-21)
>nrl3d:pir||1GPHA Glutamine phosphoribosylpyrophosphate (prp
35 p) Amidotransferase (EC 2.4.2.14), chain A - Bacillus subtil
is >nrl3d:pir||1GPHB Glutamine phosphoribosylpyrophosphate (prpp)
Amidotransferase (EC 2.4.2.14), chain B - Bacillus subtilis
>nrl3d:pir||1GPHC Glutamine phosphoribosylpyrophosphate (prpp)
Amidotransferase (EC 2.4.2.14), chain C - Bacillus subtilis
40 >nrl3d:pir||1GPHD Glutamine phosphoribosylpyrophosphate (prpp)
Amidotransferase (EC 2.4.2.14), chain D - Bacillus subtilis
atgcttaaggattcaggagctaaccgcattcacgtaagaattgcttctcccgaattcatg
ttccctagtttttatggtattgacgtatctacaacagctgaactcatctcagcaagtaag
45 tctcctgaggaaattaaaaatcatattggtgcagattctcttgcattatttaagcgttgat
ggcttaatcgagtcctataggacttgattatgatgcgccatatcatggcttggtgtagaa
agttttacaggtgattatccagcaggactttacgattatgagaaaaattataaaaagcat
ttaagtgaacgtcaaaaatcatatatagctaataataaacattattttgatagtgaggga
aatttacatgtctaa

Sequence 1274

MLKDSGANRIHVRIASPEFMFSPFYGIDVSTTAEILISASKSPEEIKNHIGADSLAYLSVD
GLIESIGLDYDAPYHGLCVESFTGPDYAGLYDYEKYKKHLSEKQSYIANNKHVFDSEG
NLHV*

Sequence 1275

Contig_0580_pos_3583_5061,
is similar to (with p-value 0.0e+00)
>sp:sp|P12048|PUR9_BACSU PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXA

MIDE FORMYLTRANSFERASE (EC 2.1.2.3) (AICAR TRANSFORMYLASE) /
 IMP CYCLOHYDROLASE (EC 3.5.4.10) (INOSINICASE) (IMP SYNTHET
 ASE) (ATIC). >pir:pir|A29183|DTBSPH purH bifunctional enzyme
 - *Bacillus subtilis* >gp:gp|J02732|BACPURF_10 *B.subtilis* pur
 operon encoding purine biosynthesis enzymes, 12 genes. NID:
 5 g143363. >gp:gp|Z99107|BSUB0004_100 *Bacillus subtilis* compl
 ete genome (section 4 of 21): from 600701 to 813890. NID: g2
 632866.

atgaaaaaagcaatattaagcgtttctaataaaagtggaattgtagagtttgcaaaagca
 10 ttaactaatttagactatgaactgtattctacgggtgtacaaaacgtgtattagaagat
 gcgaatatcaaatattaagtcctgtcagaattaacacaatttccagagattatggatggt
 cgtgttaaaacactacatccagcagtcctatggtggtatttttagctgatcgagataaagaa
 catcatttagagcaattacgagaacaacatattgatttaattgataggtagtagtcaac
 ctatatcctttccaacagactgttgctcaacctgatgtaacagaaactgatgcaatagaa
 15 aatattgatattggtggacctacaatgttaagagcagctgctaaaaactttaacatggt
 acaactatcgctacatccttccgattacaacgaggttaattgagagaattaaaaatcatcaa
 ttggacgaagcatatagaaaatcgctaattggttaagttttccaacatacaaatgaatat
 gatcattgctattgttaactatttcaagacaataaagaaacactaagatatggcgaaaat
 cctcaacaatctgcatattttgttagaacatctgatagcaaacatacgattgctggtgca
 20 aaacaattacatggtaaaacaattgagttttaataatattaagacgcagatgcagcgctc
 agtttagtaaaaaaattcaacgagccaactgctgtagcagtaaaacatatgaaccgtgt
 ggagtaggaattggacagtcgattgatgaagcatttcaacatgcatatgaagcggataat
 caatcaatatttggcggaattatagcattgaatagaacggtagatgttaaattagctgaa
 gcattacattctatctttttagaagtagttatcgcacctcaatttactgaggaagcttta
 25 aaaaatttgacacaaaagaaaaatattcgtttattacaaatagatatgacaattgataac
 gctgaacaagaatttgtttccgtttcaggtggttacttagtacaagataaagataataaa
 gatgtgactcgaaatgacatgactgttgctaccgacattcaacctacagaagcacagtgg
 gaagctatgctcctaggttgaaagttgtaagtgccgttaagagtaatgcagtgatattg
 agtaacaacaaacaaacagtttggtataggtgcagggcaaatgaatcgtgtaggttccgct
 30 aaaattgcaatcgaaagcaatagaaattaacgataatgttgcgcttgtttcagatggt
 ttcttccaattgggagatacagttgaatatgctgccgaacatggtattaaggcaattatt
 caaccaggtggttcaattaagatcaagattccattgatatggctaataaatatggcatt
 acaatggttatgacaggtatgcgtcatttttaaacattaa

35 Sequence 1276
 MKKAILSVSNSKSGIVEFAKALTNLDYELYSTGGTKRVLEDANINIKSVSELTQFPEIMDG
 RVKTLHPAVHGGILADRDKEHHLEQLREQHIDLIDMVVNLYPFQQTVAQPDVETDAIE
 NIDIGGPTMLRAAAKNFKHVTTIVHPSDYNEVIERIKNHQLDEAYRKSLMVKVFOHTNEY
 DHAIVNYFKDNKETLRYGENPQQSAYFVRTSDSKHTIAGAKQLHGKQLSFNNIKDADAAL
 40 SLVKKFNEPTAVAVKHMNPGVGIGQSIDAEFQHAYEADNQSI FGGI IALNRTVDVKLAE
 ALHSIFLEVVIAPQFTEELKILTQKKNIRLLQIDMTIDNAEQEFVSVSGGYLVQDKDNK
 DVTRNDMTVATDIQPTAEQWEAMLLGWKVVS AVKSNVILSNNKQTVGIGAGQMN RVGSA
 KIAIERAIEINDNVALVSDGFFPMGDTVEYAAEHGIKAI IQPGGSIKDQDSIDMANKYGI
 TMVMTGMRHFKH*

45 Sequence 1277
 Contig_0581_pos_7809_6784,
 putative peptide of unknown function

atgaatattgcagctgttctacaaacaattccatcactggcactacttggtttgatgata
 50 ccaattttcgaattgggagacttccggcaattatcgcttagttgtatatgcgttactt
 cagcaagatggttccagattattgacttagctcaccgtatgaaattaaatgaacctatc
 gatattactaaacgttatcatgatcgtagttttattcgttgggtacgaatcaaattcca
 gacgttgggtgataaagtagttaaaagcgctgtagctaaaggctatgatagagtatata
 caagttttggctcctatgtataaaggtaacgctggtattaagagacttaaccaagttcta
 55 caatctattcttaatccgaagcaacaagatgatcgtagaatagaatttgggtgaagctgtg
 tttagaaaaggggataaagtagttcagtttagttaatcgacctaatgataatatattta
 ggggatataggtataatagtaggtatattttgggccaaagaaaatgctctaaataaggat
 gtgttagttgtagattttgaaggtaatgaaattacatttactaaacaagatttaattgaa
 ctaacacatgcatattgtacatctatccataaatcacaaggttcagaatttcctattgta

attatgcctattgtagacaatattataggatgttacaacgtccattctttatacagga
ttaactagagctaaacaatcacttggtttgcttggtgaacaagaagcatttgatataggt
ttaaaaaaacaatggacaaatcagattaacgcaattaaatgatttgtaaaatcgtat
ggacaaaacaaagataatttaactacaaataaacaacgattaacgaacaaaaagaaat
5 aacaatcatctggatttgaaaaatgaaaaagaaatgatatccaattaaacgagtcgaca
attttccaaatcgatccaatgattaatatgggggaaatgacgccatagacttcggtgaa
cgttga

Sequence 1278

10 MNIAAVLQTIPTSLALLGLMIPIFGIGRLPAIIALVVYALLQQDGSIIIDLAHRMKLNEPI
DITKRYHDSFIRCGTNQIPDVVDKVVKS AVAKGYDMSDIQVLAPMYKGNAGIKRLNQVL
QSILNPKQQDDREIEFGAEVFRKGDVQLVNRPNNDNIFNGDIGIIVGIFWAKENALNKD
VLVVD FEGNEITFTKQDLMELTHAYCTSIHKSQGSEFPVIMPIVRQYYRMLQRPILYTG
LTRAKQSLVLLGEQEAFDIGLKTNGQIRLTQLNDLLKSYFGQNKDNLTTNKQTINEQKEN
15 NNHLDLKNKEKENDIQLNESTIFQIDPMINMGEMTPYDFVER*

Sequence 1279

Contig_0581_pos_3057_2749,
putative peptide of unknown function

20 atgacagaacataatcacgatgctgagttacaataaataatgaagaggaattacttacg
ttatatgatgaaaacggaatgaagttttataccgtaaaatgttagaattttatcatcca
gaattcaaaaaagaatatgtcgttcttgacagaagaaggtgcacaatcagatgacgaagat
atgattgaacttgtaaccaatgataaatgaacctgatgagtcgtggtgatggtgggaaatta
gtccctattgaaacagatgaagaatgggatatgattgaagaagttgtaatactgagatt
25 aacgaataa

Sequence 1280

MTEHNHDAELTINNEEELTLYDENGNEVLYRKMLEFYHPEFKKEYVVLAEEGAQSDDED
MIELVPMINEPDESGDGGKLVPIETDEEWDMIIEVVNTEINE*

30

Sequence 1281

Contig_0581_pos_1682_759,
putative peptide of unknown function

atggttgattatttagttactcctaagtcataactcatatggaaacactaatagataaa
35 ggtgcagacgcatttggtatttggtgaacaaaatttggtttaagactgccgggagaattt
aatcgtgatgctatgcaagaagctgtagcatttagccataaaaaatacaaaaaagataac
gctgctgtgaatggtattttccataattaccacttagatgccttgaagactatattaac
ttttacatgatattcaagtagatcgcatatatatttggtgatccagctgtcgttatgtat
gttaacaacacagcagcatccaattccattaaattgggatgctgaaactcttgtaacgaat
40 tattttcagtgtaattactgggggaaaaaagggtgcaaatagagcagttttagctcgagaa
cttagtttagatgaaataattcatattaaagagcatgctgatgtagagatagaagttcaa
gttcatgggtatgacttgatgtttcaatccaaaagaatgctattaggaaattattatact
ttccaagagcgacaaatgaagatagaacgccaacatgattatggagacttattattatat
gatgaagaaagagataataaatatccagttttgaagattataatggtactcacatcatg
45 tctccaaatgatattctgtttaatagaagaattagagccttttttgaggcaggaatagat
gcgtttaagatagatggtattttacaaagtgaagaatatataaatgtagtcacagagcaa
tatcgagaagctatagattttatttaaatgaagatccggatgcatatgaagatgaaaatcc
atgctcgttgatcctatagaagaataacaacctgaacatcgctccattcgacgaaggtttc
ttgtataaacaacagtatatataa

50

Sequence 1282

MVELLVTPKSIITHMETLIDKGADAFVIGE QKFLRLPGEFNRDAMQEAVALAHKNNKVVY
AAVNGIFHNYHLDALEDYINFLHDIQVDRIIFGDPVVMYVVKQHEHPIPLNWD AETLVN
YFQCNYWGKKGANRAVLARELSLDEIIHIKEHADVEIEVQVHGMTCMFQSKRMLLGNYT
55 FQERQM KIERQH DYGD LLLYDEERDNKYPVFEDYNGTHIMSPNDICLIEELEPF FEAGID
AFKIDGILQSE EYINVVTEQYREAI DLFNEDPDAYEDEKFM LVDPIEEI IQPEHRPFDEGF
LYKQTVY*

Sequence 1283

Contig_0581_pos_0_741,

putative peptide of unknown function

atgaaaacttttagaagaagttaaatccaaaacccaaaagataatgaaaaagcctgaacta
 cttgcaccagctggtaacttagagaaaacttaagattgctgttcattatgggtgcagatgca
 5 gtgttttttagtggtgccaagaatatggattacgttctaatactgataattttactatggaa
 gaaattgctgaaggtgtagactttgctaactggttatggcgctaaaatttatgttacaacg
 aatattattgcacatgatgaaaatatggaggactagaagagtacttacaaaaccttgaa
 tctacaggtgctactggtatcatagtagcggtacctcttatcatagaaacttgtaaaaaa
 gttgcgcccagattagaaattcatttatcaacacaacaatcactttcaaattataaagct
 10 gttgaatttttgaaacaagaaggattagaccgtgttgacttgacgtgaaactggtgca
 atggaaatgagtgaatgaaagaaaaagttgatattgaaatagaagcgtttatacatggc
 gcaatgtgtatcgcatattctggtcggtgtactttaagtaacatagacagctcgagat
 tctaactcgaggtggttgtgtcgaggtgtcggtgggactatgatttacttgaagttgat
 agtgatggagaattagatttatattacgacaatagtgtgttactccttttgcaatgagt
 15 cctaaagattttaaaattaata

Sequence 1284

MKTLEEVKSKTKIMKKPELLAPAGNLEKLKIAVHYGADAVFLGGQEYGLRSNADNFTME
 ETAEGVDFANRYGAKIYVTTNIIAHDENMEGLEEYLQNLSTGATGIIIVADPLIIETCKK
 20 VAPRLEIHLSTQQSLSNYKAVEFWKQEGLDREVVLARETGAMEMSEMKEKVDIEIEAFIHG
 AMCIAYSGRCTLSNHTARDSNRGGCCQSCRWDYDLLEVDSDGELDLYYDNSDVTFFAMS
 PKDLKLI

Sequence 1285

25 Contig_0583_pos_4830_3856,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P50915|HEM2_STAAU DELTA-AMINOLEVULINIC ACID DEHYDRATA
 SE (EC 4.2.1.24) (PORPHOBILINOGEN SYNTHASE) (ALAD) (ALADH).
 >gp:gp|U89396|SAU89396_3 Staphylococcus aureus hemCDBL gene
 30 cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
 III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
 : g2589180.
 atgaaatttgatagacatagaagattgcgttcatctaagacaatgcgtgatttagtaaga
 35 gaaactcatgttagaaaagaagatttaatatatccaatatttgtagttgagcaagatgat
 ataaaaagtgaattaaatcactaccagcatataccaaattagtttaaatatttgc
 gaagagattaaagagcatatgatttaggtattagagcaatcatgttcttcggtgtgcca
 aatgacaaagacgacattggatctggtgcatatgatcataatggagttgttcaagaagcg
 acacgaatatctaagaatttatataaggatttacttattgttgcagatacttgccttgc
 40 gaatacacagaccagcgacactgtggcgttattgacgatcatacgcgatgatgtagacaat
 gataaatcacttccattacttgtaaaaacagctatttctcaagttgaagctggagctgac
 atcattgctccaagtaatatgatggatgggttttgttgcgtgaaattcgtgaaggccttgat
 caagcgggatatacaaatattcctatcatgagttatggtattaaatgatcatcaagcttt
 ttcggtccattcagagatgctgcagattcagcaccttcttttggggatagaaaaacctat
 45 caaatggatcctgcaaaccgttttagaggcattaagagaattggaaagtgatcttaagaa
 ggttgcgatatgatagattaaaccatctttaagttatctagatattattagagatgta
 aaaaaataatagcaacgtgcccagtcgtagcatataacgttagtggaagaatagtagatgaca
 aaagcagcagcgttaaatggttggatagatgaagagaaaattgttatggaacaaatgata
 tctatgaaacgtgcaggtgctgatttaataattacttattttgcaaaagatatctgtcgt
 50 tatttagataaataag

Sequence 1286

MKFDRHRLRSSKTMRDLVRETHVRKEDLIYPIFVVEQDDIKSEIKSLPGIYQISLNLH
 EEIKEAYDLGIRAIMFFGVPNDKDDIGSGAYDHNGVVQEATRISKNLKDLLIVADTCLC
 55 EYTDHGHGCGVIDDHTDNDKSLPLLVKTAISQVEAGADI IAPSNMMDGFVAEIREGLD
 QAGYQNIPIMSYGIKYASSFFGPFDAADSAPSGDRKTYQMDPANRLEALRELES DLKE
 GCDMMIVKPSLSYLDIIRDVKNNTPVVAYNVSGEYSMTKAAALNGWIDEKIVMEQMI
 SMKRAADLIITYFAKDICYLDK*

Sequence 1287

Contig_0583_pos_3805_2549,

is similar to (with p-value 0.0e+00)

>gp:gp|U89396|SAU89396_4 Staphylococcus aureus hemCDBL gene
 5 cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
 III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
 : g2589180.

atggagcaagctgagaaattaatgcctggcggtgttaacagtcgccgtaagagcatttaaa
 10 tcagtagacacaccagctatttttatggatcatggtgaaggatctaaaatatatgatatt
 gatggaatgaatacattgattatgtgctaagttggggccatttaattctgggacataaa
 aatcaacaagttatatccaaattacatgaagcagtagataaagggtacaagcttcggcgct
 tcaacacttcaagaaaataaacttgctgaacttggtgattgaccgtgtaccttcaattgaa
 aaagtaagaatggtttctcaggaactgaagctacttttagacacacttcgtttagctagg
 15 gggtatacaggacgtaataaaaattataaaattgaaggggtgttatcatggacacagtgat
 tctttattgattaaagcaggatcagggtgttgcaacactagggtttacctgattcaccaggc
 gtccttgaaggtattgctaaaaacactatcacgggtgccatataatgatttagattcactt
 aatatactgctgcgaaagattggcgatgatattgctgggtgttattgttgaaccgggtgct
 ggaaatatgggtgtagtgcctccagtgaaatggatttctacaagggttaagagatattact
 20 aatgaatatggagcattacttatatttgatgaagtgatgactggtttccgtgtaggttat
 aattgtgcgcaaggatactttggtgtaacacctgatttaacttgcttaggaaaagtgata
 ggtggaggtttaccggttgagcttttgggtggtaaaaaagaaattatggattacattgct
 cctgttgggactatttatcaagctggcacactttcaggtaatccttttagcaatgactagt
 ggttatgaacattgagtcacttactcctgaatcttatgagtattttaattctctagga
 25 gatatacttgaaaaagattaaaaagaggtatttgctaagcataatgttccaatcacagta
 aatcgcgctgggttcaatgattgggttacttcttaaatgaggggcctgtaacaaatttgag
 gaagcaataaaaagtgatttaaaattatttagtaatatgtatagagaaatggctaaggaa
 ggtgttttctaccaccttcacaatttgaaggacatttttatcaactgcacatactaaa
 gatgatattgagaaaaactatccaagcatttgataatgcattaagtcgtattgtgtga

Sequence 1288

MEQAEKLMPPGVNSPVRFAKSVDPAPAFMDHGEKSKIYDIDGNEYIDYVLSWGPLILGHK
 NQQVISKLEAVDKGTSFGASTLQENKLAELVIDRVPSIEKVRMVSSGTEATLDTLRLAR
 GYTGRNKIKFEGCYHGHSDSLIKAGSGVATLGLPDSGPVPEGIAKNTITVPYNDLDSL
 35 KLAFEKYGGDDIAGVIVEPVAGNMGVVPPVNGFLQGLRDTNEYGALLIFDEVMTGFRVGY
 NCAQGYFVGTPDLTCLGKVIGGGLPVGAFFGGKKEIMDYIAPVGTIYQAGTLSGNPLAMTS
 GYETLSQLTPESYEYFNSLGDILEKGLKEVFAKHNPITVNRAGSMIGYFLNEGPVTNFE
 EANKSDLKLFNSMYREMAKEGVFLPPSQFEGTFLSTAHTKDDIEKTIQAFDNALSRIV*

Sequence 1289

Contig_0583_pos_2152_1133,

putative peptide of unknown function

atgtaaagctctgttattaaaaatgttacatgtgattttgccatttatgtttggaccaata
 tttagcggcgttattatgtgtataaagtattaaaaataaaaatacagatggccattttggttg
 45 agtcaaatgggtttaaactacttggagttcaaattggctctaccttcacacaacaagtga
 attaaaagacataagtaaaaaattggctaactatcgtttttgtcactatcctactaatttta
 tttagctttgataaattgcattcttttttaagaaaattgcacaagtaaattagaaactgca
 atttttaagtgttataccaggtgcgctaagccaaatgttagtgatggcagaagaaaataag
 aaagcaaatatattagttgtgagtttaacacagacatcacgtgtaaatattgtgttatt
 50 ttagtaccacttatttcgtatttttttcaggataaccatcatgaaatgaatcatactaca
 atggaagtacccacactttctcagactttaaatatatggcaataatcatcttattctca
 atgggtgggaatcatctatataggaatgtcaaaaataaactccccactaaacaattatta
 gcacctataatagtttttaattatatggaatatgacaacacatttaacattttcactagat
 cattgggtgttagccacagcgcaacttatttatatgatacgtattggattacagattgcc
 55 aacttaatgagtgatttaaaagggaagaattgcaatagcaatagcctttcaaaatataatg
 ctcatagtcacacggtttataatgataataggaatacatttgattactaatgaatccatc
 aatgaattgttttttaggagcagcaccaggaggtatgagtc aaatagtttttagtggtatg
 gctactggagctgatgtagcgatgatttcaagctatcacatttttagaataattttttata
 ttatttgcattgcgccactaatgggtatttttattaatgttaaataaataataaatga

Sequence 1290

5 MLSSLLKMLHVILPFMFPGPILAAALLCVKVLKLRWPFWLSQIGLILLGVQIGSTFTQQV
 IKDISKNWLTIVFVTILLILLALIIAFFFKKIAQVNLETAIVPGALSQMLVMAEENK
 KANILVVSILTQTSRVIFVVLVPLISYFFQDNHHEMNHTTMEVPTLSQTLNIWQIIILFS
 MVGIIYIGMSKINFPTKQLLAPIIVLIIWNMTHTLTFSLDHLLATAQLIYMIRIGLQIA
 NLMSDLKGRIAIAIAFQNI MLIVTTFIMIIGIHLITNESINELFLGAAPGMSQIVLVAM
 10 ATGADVAMISSYHIFRIFFILFVIAPLIGYFINVKLNK*

Sequence 1291

Contig_0584_pos_2306_3883,
 is similar to (with p-value 3.0e-72)
 >sp:sp|P47994|SECA STACA PREPROTEIN TRANSLOCASE SECA SUBUNIT
 15 . >pir:pir|S47149|S47149 secA protein - Staphylococcus carno
 sus >gp:gp|X79725|SCSECA_2 S.carnosus (TM300) secA gene. NID
 : g499333.
 gtgattcaaaatgaacggcaaatgcttaagtgtttaaaaaatgtacatggtgctgcaaac
 gagtatcaaggatgaactgatattttgcataaagtcaaagatggcgctgtggatgatagc
 20 tatggattcaagtggcaaaattagcggatttacctaataagtcattgatagagcgcaa
 gttatattaaatgcatttgagcaaaaaccttcgtatcaactctctcatgagaatactgac
 gatcaacaaacggttcgctcgatataacgattttggtcgaacagaagaagagcaatcagtt
 atagaacacatacatcaaatcataattatattttgatggtgagattgtgcttatagat
 agaataactggtcgatgtacctggaacaaagcttcagtcctggtttacatcaagctata
 25 gaggtctctggaaaatgttgaaatttctcaagatatgagtgtgatggcaaccataaacattc
 caaaacttatttaagcaatttgatgaattttcaggtatgactggaacaggtaaattagg
 gaaaaagaattctttgatattatcaaaagtgttatagagattccgactcacagtcg
 attgaacgagatgatagacctgatagagtatttgctaattggtgacaaaaagaacgatgca
 attttaagacagtgattggtatacatgaaactcaacaacctgtgttactaattacacgt
 30 actgcagaagcggcagaatatttttcagctgagttatttaaacgtgatataccaacaat
 ttattaatcgctcaaaatgtagctaaagaggcacaatgattgctgaggcgggacaatta
 tctgcagttactgttgctacaagtatggcagggcggtggaactgatataaagttatcaaaa
 gaggttcatgatatcggtggcttagcagtgattattaatgaacatatggataatagccgt
 gttgatcgctcaattaagaggacgctcaggtcgccaaggagatcctggatattcacagatt
 35 tttgtatcacttgatgatgatttagtaaaacgttgagtaactctaacttggcagaaaat
 aaaaacctccaaacgatggatgcatctaaactagaaagtagtgactctttaaaaaacgt
 gtaaagtcaattgttaataaagcgcaacgtgtatctgaagagactgctatgaaaaataga
 gaaatggcaaatgaattcgaaaaaagtatttagtgttcaacgagataaaaatttatgctgaa
 cgtaatcacatacttgaagcaagcgattttgatgattttaattttgaacagcttgcaaga
 40 gatgtgtttacaaaagacgttaaaaatcttgacttaagtagtgaaacgtgcacttgtgaat
 tatatatcgaaaacttaagttttgtcttcgatgaagatgtatcaaatattaatatgcaa
 aatgatgaagaaatcatacaattcttaataacaacatttactcaacaatttaacaatcgt
 ttagaagttgctgctgattcatatttaaaacttgttattgaaaatcttaataaagttaaa
 45 cacatcggttaaagaataa

Sequence 1292

VIQNERQMLKCLKNVHVAANEYQGELIFLHKVKGAVDDSYGIQVAKLADLPNEVIDRAQ
 VILNAFEQKPSYQLSHENTDDQQTVPYNDGFRTEEEQSVIETHTSNHNIFDGEIVLID
 RITGRMLPGTKLQSGHLQAIEALENVEISQDMSVMATITFQNLFKQFDEFSGMTGTGKLG
 50 EKEFFDLYSKVVEIPTHSPIERDDRPDRVFANGDKKNDAILKTVIGIHETQQPVLLITR
 TAEAAEFYFAELFKRDI PNNLLIAQNVAKEAQMI AEAGQLSAVTVATSMAGRGTDIKLSK
 EVHDIGGLAVIINEHMDNSRVDRQLRGRSGRQDGPYSQIFVSLDDDLVKRWSNSNLAEN
 KNLQTM DASKLESSALFKRVKSIVNKAQRVSEETAMKNREMANEFKSISVQRDKIYAE
 RNHILEASDFDDFNFEQLARDVFTKDVKNLDLSSERALVNYIYENLSFVFEDEVSNINMQ
 55 NDEEIIQFLIQQFTQQFNRRLEVAADSYLKLVIENLNKVKHIVKE*

Sequence 1293

Contig_0584_pos_4844_6181,
 putative peptide of unknown function

atgaaagtcaaaagtatttcacggttcttttcaatgaagaaagtgacgctaagtttcggt
 actttattttattggagtagggacaataggttcatacaatcagtttgctgatgcaagtacg
 aaaacgcaacaaacacatgtaactaagacatctccaactcaaaagacgacgtccaatttt
 aaacggttcagttaaagatacgtctgtttaaacttagagctacatcaacaaaagagctaca
 5 tcaacaaaacgagctatatcaccocaaacatcatcaactaaaaaaactacaatagcaaaa
 aaatctaccacagtaataaaaacgcgcacacaaccaggactcagcctaccattcgtaag
 agttcaacaacttcaacacggttcaaaaacaatacctacttctgtgaaacgcacaacttct
 cataaagcaactactgtgtcgccaacttctaaagctaaaatatcaacaaagacacaacaa
 tcaactaaaagtcatacaacttcagtttaagaaaaacactacacaactaagtaaaacaaaa
 10 tctccgtcaacgtcaacaaaatctaaaacagttcaatcctctacgacaaaggcacaacct
 actttatcgactcaagtttagtacaactactaaagcaaaagcaactttcaacgccaactact
 tctaaaacacttgtagcagtaaaagcttttagtaagtttagcatctacagaacgtaaaatagat
 aaataccaatcgatgactcagtttagaaaaagaacaaactgaaggtgtagattggagaaaa
 gatacaaaaaaacaggggaatcaagtaactcattgtggctccacatggcggaagtattgaa
 15 caaggtacaacagaattaactaaagcatttagcagataaaaggaattatgattattattca
 tttgaaggtattcgacctaataaataactctgaattacatgtgacgtctacacattatgat
 gatccgacattaaatcaaatgattaaaaaccgtactgcaactatttcgattcatggcgca
 tcaggtactgaggagattatctatcttgggtgggcccgttcagatttaagaaatgctata
 gagaacaaacttgtaggagcgtggatttacagtttaaagttccaccagagtatctagggtgt
 20 caaaataataaaaaacttcattaataaagaagacaataacactggcggttcagttagaatta
 acgactgctttaagaaaaagcattcttttaaaaatggagataactagtacaaaaaatcgtaac
 aataaagaaaattggacaccaacaatggaagcattttattaatgcatttatgaaggtatc
 aatcaaacgtattcataa

25 Sequence 1294
 MKVKSISRFFSMKKVTLFSVTLFIGVGTIGSYNQFADASTKTQQTHVTKTSPTQKTTSNF
 KRSVKDTSVKSRATSTKRATSTKRAISPKTSSTKKTIIAKKSTTVNKRTRTTTTTQPTIRK
 SSTTSTRSKTIPTSVKRTTSHKATTVSPTS KAKISTKTQQSTKSHTTSVKKNNTQLSKTK
 SPSTSTKSKTVQSSTTKAQPTLSTQVSTTTKAKQLSTPTTSKTDSSKALVSLASTERKID
 30 KYQSMQTLEKETTEGVDRKDKNTGNQVLIVAPHGGSIEQGTTELTKALADKGNYYYS
 FEGIRPKNNSELHVTSTHYDDPTLNQMIKNRTATISIHGASGTEEIIYLGGPRSDLRNAI
 EKQLVGRGFTVKVPPEYLGQNNKNFINKEDNNTGVQLELTALRKAFFKNGDTST'KNRT
 NKENWTPTEAFINALYEGINQTYST*

35 Sequence 1295
 Contig_0584_pos_6957_9221,
 is similar to (with p-value 4.0e-31)
 >sp:sp|P13485|TAGF_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN
 F. >pir:pir|S06049|S06049 rodC protein - Bacillus subtilis >
 40 gp:gp|X15200|BSRODC 2 Bacillus subtilis rodC operon. NID: g4
 0098. >gp:gp|Z99122|BSUB0019_69 Bacillus subtilis complete g
 enome (section 19 of 21): from 3597091 to 3809700. NID: g263
 6029.

gtgaaaagcttatttagaaattggacatgaggtacactactttaattatcaggactataat
 45 aaaagtgatatcacaaaactaattattttacgaaggtttgagcacaagcatcttcatatt
 catcaatttaagtggaaaagaacttgctcatggagacctacttataattactagagaa
 accttttttaatacatgcatatctagttaaaaaattaaatagcaagattaagattgttggt
 gaaatacatgggtccattggaatatattaatgagaatatagatttagcattagactgtatt
 gattgtgttcgagtgagtacagctagaattaaaaatgaatttatagctaaatatgactat
 50 catcgggttttttaatacaatacgtaaatgcacaacatatcgatttaaaatcagagccgata
 aataactaaacgaaattttttaattaaagcacgttttgaggatgaagttaaagatatcca
 tatattattaaattgtttaattacatcattaaaaaccaaattgttgatgatgctcaactt
 tatttaaatgagatatggtccttcagaaatgctttacaaaaatttgataaattactatcat
 cttaatgattatattcatattaatgaaaaagaaccacagagttatatttatgtatctagt
 55 tcgccatatgaaacgctcggtatttctatattagaaacgattgcacagggttaataagct
 ctagtttactatggtgatgataacgtgttaaaggatatctatgcaccatatgaagcgata
 cgttttttaacaaagatatgattaaagatagtaaaataattaaagactttctaaactat
 aaatatagtcactgtgatcgacaaaaagattatcgacagttgaaaagtacgtttaaatgc
 attaatatggacaggaatttttaataatgttgaaactttctcttcatctcaacatgta

aaagtgaagaaaattcatcgacatctcggtagtgaaaaacaaatagatatagcaagtcgt
 ttaaaagagagtcggttgatgaatttaattagaaaaataaatattttatataaatgt
 aaaacgtactatgaaaaaagaacgcataatgagctatatcaaaaatttaaatcaaatacct
 gtagacgacgattcaatatttatagagtccttccatggcaagaatttttagtgagatcct
 5 aaatatattgctcttgctattaagagacagtatgatcataaaaaatatatgtgagttca
 accaattcacttggtgatatggaaatcaaacggttacggttttacacctgttcgatttgga
 agcgagaaatatattaaaacggttagaaagtgaagtatgtttttatcaatggtaactcg
 tgggataaaagtgtacaagtcctcagatcagatatttattcaaacatggcacggttttcca
 ttaagaaaaatggtaatgatttaaatgaacaacatgaaagacaacaacaactagaggca
 10 ttcataccacgcataaaaaatgggattacattttgacatcatcagatattaatacgcag
 ttggttggaatctgcttttatgctaataaaaaatccaaatcttaaaagtctagaatcagcg
 gcacctagaatgaatatttaataaataataaatttacaagagcgccagcagttacag
 cttaaatatatgtataagatagatgatgataaaaaatatattatattgtccacttg
 agggaaaaatcaaagaaaagaagtcactcagattaatttaaaagatttacttaaatattta
 15 ccagagaattatgagattattgtgaaacttcactcctaatagaagtcatttaagaaccaga
 tataatcaaatagataatcgaattcactgattttcaatgaacttggtgatattcaagaa
 ctgtatattctgagtgatgtatgattacagattactcgctcgaccatttttgactatata
 catttaaacagccagtcctttattcttcaagaagcagagcaacaataaaacaaagtgtt
 ggtttttatttttgattgtttgaagtgggtgattttcttaaagcctctttaaataaacgc
 20 atgttagctaaacaaattttagcactgattatataaattattcaaaagtgggttcacgt
 ttgatgaaacaagatagttcgaaaagcagtgaaaagttaatggccgaaattcttggggaa
 ccagaatatccaagttcatcaaactgcaaacacagatttcttaa

Sequence 1296

25 VKSLLEIGHVEHYFNYQDYNKSDITKLIIEGLSTKHLHIHQFNSGKELAHGDLIIITRE
 TFFNHAYLVKKLNSKIKIVGEIHGPLEYINENIDLALDCIDVRVSTARIKNEFIKYDY
 HRVFNQYVNAQHIDLKSEPINTKRNFLIKARFEDEVKDISYIIKLFNYIIKNQIVDDAQL
 YLIGYGPSEMLYKNLINYYHLNDYIHINEKEPQSYIYVSSSPYETLGYSILETIAQGNKA
 LVYGGDDNVLKDIYAPYEAIRFLTKDMIKDSKIIKDFLNKYSHCDRQKDYRQLKSTFKC
 30 INYQEFLLNNVETFFSSQHVKKIHRHLGSEKQIDIASRLKESRWMNLIRKNKYLNFNC
 KTYYEKRTHMSYIKNLNQIPVDDDSIFIESFHGKNFSGDPKYIALAIKRQYDHHKIYVSS
 TNSLVDMEIKRYGFTFVRFGSEKYIKTFRKCKYVFINGNSWDKVYKSSDQIFIQTWHGFP
 LKKMVNDLNEQHERQQQLEAFIPRMKKWDYILTSSDINTLLESFMLNKNPNLKVLEYG
 APKNEYLINNNNLQERQQQLKMYKIDDDKKYILYCPTWRENQRKEVTQINLKDLLKYL
 35 PENYEIIVKLHPNESHRLTRYNQIDNRIHCYFNLVDIQELYILSECMITDYSSTIFDYI
 HLNKPVFILQEDQYKQSVGFYFDLFEVGFELKASLNERMLAKQICSTDYINYSKVVR
 LMKQDSSKSSEKLMAEILGEPEYPSSSNCKQQIS*

Sequence 1297

40 Contig_0584_pos_13726_14622,
 putative peptide of unknown function
 atgaagtttgcatatattcaatcgattcgtaatgagatttcaattattttaataattcta
 ttattttttgcgcttatattttatgtgttttctttaccttttgatgcatacgtactagca
 atcagtataatattactattgatgtgtgtacgttggtggataaaagtatttaagtttttaa
 45 aagaatgaacatcttaaagataaaagtacatatttagaacatgagttagcacatgttaag
 aatcagcaaatgaatcgcgaacgatgttgaaagtatttttttaacatgggtacatcaa
 attaaaacacctatcactgcctcacaattacttttgagagaaacgaggagaatgtagtt
 aatcgtgttcgacaagaaattgtgcacattgataattatacaagtcctcgcatgaattat
 ttaaaattattaaatgaagagtcagatatgacaattaccaaagtgcaggttgatgatttg
 50 attcgcccggtgatttttaaaatatagaattcagtttattgaacaaaagacgcaaatccat
 tatgaaaaaagtggagacattattttaaccgatgcacaatgggcttctataatgatagag
 caacttttaaatgcttttaaaatatgctaaggtaaagatatgttgatagattttgat
 gttgccaatcaaactctacagattaaagataatgggtattgggattagtaaagcagattt
 cctaaaatttttgataaaggatactcaggatttaacggtagattgaatgaacaatcaact
 55 ggtataggtctatttatagtgcaacacattgcaaatcatttaaatatacaagtaactgta
 caatcagagttgaatcatgggacagatttttttatacattttactaaagaaaaataa

Sequence 1298

MKFAYIQSIRNEISIIILLFFALIFYVFSLPFDAYVLAIISIIILLMCVRWWIKYLSFK

KNEHLKDKVAYLEHELAHVKNQQIEYRNDVESYFLTWVHQIKTPITASQLLLERNEENVV
 NRVQRQEIIVHIDNYTSLALSYLKLLNEESDMTITKVTVDLIRPLILKYRIQFIEQKTOIH
 YEKSEDIILTDAQWASIMIEQLLNALKYAKGKDIWIDFDVANQTLQIKDNGIGISKADI
 PKIFDKGYSGFNGRLNEQSTGIGLFIQHIANHLNIQVTVQSELNHGTVFFIHFTEK*

5

Sequence 1299

Contig_0584_pos_14789_0,

is similar to (with p-value 5.0e-41)

>sp:sp|P42423|YXDL_BACSU HYPOTHETICAL ABC TRANSPORTER ATP-BI
 10 NDING PROTEIN IN IDH 3'REGION. >gp:gp|D14399|BACIOLO_13 Baci
 llus subtilis 15 kb chromosome segment contains the iol oper
 on. NID: g709980. >gp:gp|Z99124|BSUB0021_68 Bacillus subtili
 s complete genome (section 21 of 21): from 3999281 to 421481
 4. NID: g2636442. >gp:gp|D45912|D45912_2 Bacillus subtilis g
 15 enome sequence between the iol and hut operon, partial and c
 omplete cds. NID: g1408482.

atggctttaaatcaaatgaaccttgaaattgatgaaaatgaattttagcaatcatgggc
 gagtcagggttcaggtaaatctacattactcaatttaattgctacttttgatcgtacaact
 gaaggattataaaagtttagacgagttgcccgttaataatgaagaataaagacattgca
 20 cgcttttcgcagagaaatgatgggatttgtgtttcaagattttaatgtgttgatacgatg
 tcgaacaaagataatattttgatgcctcttgacttgcaaatgaacgtccgaaaataatg
 caaaaacgcttaataagtaagtaacaatttaggaattgaagacttgcttgaaaaatat
 ccgtctgaaatatctggggacaaaaacaacggatagctatagcccgtgcgttgatagca
 cgacctaaattattattagctgatgaaccactgggtgcacttgattcaaaaacctctaaa
 25 aaccttatgtgtttatttcgaaaaattaatacaaaagcatcaactatattaatggtgaca
 cattcaaatattgacgcgtcatatgcgaaccg

Sequence 1300

MALNQMNLEIDENEFAIMGESGSGKSTLLNLIATFDRTEGLIKLDELPLNQLKNKDIA
 30 RFRREMMGVFVQDFNVLTMSNKNILMPLVLANERPKIMQKRLMEISEQLGIEDLLEKY
 PSEISGGQKQRIAIARALIARPKLLLADEPTGALDSKTSKNLMCLFRKINQKHQITLMVT
 HSNIDASYANR

Sequence 1301

35 Contig_0584_pos_12548_11079,

is similar to (with p-value 0.0e+00)

>sp:sp|P19405|PPB3_BACSU ALKALINE PHOSPHATASE III PRECURSOR
 (EC 3.1.3.1) (APASE III). >pir:pir|B39096|B39096 alkaline ph
 40 osphatase (EC 3.1.3.1) III precursor - Bacillus subtilis

atgaaatttatgaacaaaatgggtaagacgacgcttgctcatcaatcgtagcagcatcc
 gttttaagtacggtaaacgtatcatatgcttcaggtagctcagaacaaagtgtcaaaact
 aagcaaacacaaaacgatgccattgctttcggaacacaaaaaatccaaaaatgtcatc
 ttcattggttgccgatggtatgggaccttcttttaacactgcatatcgttattataaaaat
 aagcctgggtgctaagaaaatgactccaactgcattcgataaatatctaaaagggaacaaat
 45 cgtacttattctaatgatccctaaagaaaatgttacagactctgctgctggaggaacagct
 ttttagtaccggtcacaaaacatataaacgggtgcgattagtgttgatacaataaaaaacca
 attaaatctgtgctagaacaagctaaagaacaaggaaaatcaactggttagtaactact
 gctgaacttactgatgcaacacctgctgtatatgctgctcatgtagattcacgcgacaaa
 aaagatgaaattgcacagcaattttataatgataaaataaatggtaaacataaagtcgat
 50 gtgatgttaggtggcggtgcaaaaatacttcggtaagaaaaataaaaatttagcgaaaaaa
 ttcaaaaaagatggttatgatatcgtttctaataaagatgaattaaatcaatcacaaagc
 aagcaagtttttaggtactttctcagaaaaagatatgccattacaaatagatgcacctcaa
 tctaactccggttgctagtagacatgcaaacagtgactaaataaattaagtaaaaaataat
 aaaggattcttttaattgggtgaggtgcttcaattgataaagctgccaccctaatgat
 55 atcactgggtgtgatgtctgaaatgtctgggttcgaaactgcttttgataatgctattaat
 tatgcaaagacacataaagatacacttgtttagcaactgcagaccactcaactggcggt
 ctatcaaccgcaaaaggtaagattataaatggaatccagaggctattcacaagatgaaa
 cattctggaatgtatatgacaaaacaaatcgctgatggaaaagatcctgaaaaagtaatt
 aaagatggatacggattgatttcccaataaacaactcgataaagtcaaaaaagcagca

gacgagcttcacaaattacaaaaagaaggtaaagatgacaaagacgaaaaagttgtagaa
 caaacaacaaaattacaaaatgcaattcaaaaaccaattaacgatgcttcacacacaggt
 tggacaaccaatggccatacaggtgtagatgttaacacatatgcatatgggccaggttct
 aacaaattcaaaaggaatatggaaaatacccaaagcgtaaaaaacttatttgactttttc
 5 ggaacaatgtaacatcaaatcaaaattaa

Sequence 1302

MKFMNKMGTTLASSIVAASVLSTVNVSYASGSSEQSAQTKQTQNDIAIFGNTKNPKNVI
 FMVGDGMGPSFNTAYRYYKNKPGAKKMTPTAFDKYLKGTNRITYSNDPKNVTD SAAGGTA
 10 FSTGHKTYNGAISVDTNKKPIKSVLEQAKEQKSTGLVTTAELTDATPAVYAAHVDSRDK
 KDEIAQQFYNDKINGKHKVDVMLGGGAKYFGKENKNLAKKFKKDG YDIVSNKDELNQSQS
 KQVLGTFSEKDMPLQIDAPQSNPLLVDMQNSALNKL SKNNKGFFLMVEGASIDKAAHPND
 ITGVMSEMSGFETAFDNAINYAKTHKDTLVVATADHSTGGLSTAKGKDYKWNPEAIHKMK
 HSGMYMTKQIADGKDPEKVIKDG YGIDFPNKQLDKVKKAADELHKLQKEGKDDKDEK VVE
 15 QTTKLQNAIQKPINDASHTGWTNGHTGVDVNTYAYGPGSNKFKGNMENTQSAKNLFDF
 GNNVTSNQN*

Sequence 1303

Contig_0584_pos_2285_1518,
 20 is similar to (with p-value 2.0e-22)
 >sp:sp|P09122|DP3X_BACSU DNA POLYMERASE III SUBUNITS GAMMA A
 ND TAU (EC 2.7.7.7).
 atgtttatacaggatgatagatatcatcaatgatacactagtatccattaggttcagtgt
 aatcaaaagtgttcattttgaagtgttgctagttaaacttgcagaaatgattaagacacag
 25 cctcaaaactgtacaaaatgtagcaacagcatcggtagctaatagaaccagataatgagatg
 ttattacaacgttttagaacaacttgaaaatgagcttaaaaccttaaaagaacaaggatc
 aaaactaataaaagttagtcaacaacctaaagaaccaacacgtacgattcaacgatctaaa
 aatacgttttctatgcaacaaatagcgaaagtattagacaaagcaacaaagatgatatc
 aaattgttgaagaaccattggcaagaagtgttgatcatgcaaaaagtaataataaaaag
 30 tcttttagtaagtttgctactgaattcagaaccagtagcagctagtgaagatcatgtgtta
 gttaaatttgatgaagaaattcattgtgaaatagtaaataaagatgatgaaaagagaac
 aatattgaaagtgtagtttgtaatatagtttaataaaactgtcaaagtagttggagtgcg
 gctgaccaatggctgagagtgagagcagagtacttacaaaatcgtaacaccaatgaaaca
 catcaaaagcgaaaaacaaagcacacaacagtctcaacaaatagatatgtctcaaaaagct
 35 aaagacttattttggtgaggaaactgtacacttagttgatgaagactga

Sequence 1304

MLYRMIDIINDTVSIRFSVNSQSVHFVLLVLAEMIKTQPQTQNVATASVANEPDNEM
 LLQRLQLENELKTLKEQGIKTNKVSQPKKPTRTIQRSKNTFMQQIAKVLDKANKDDI
 40 KLLKNHWQEVIDHAKSNDKSLVSLLLNSEPVAASEDHVLVKFDEEIHCEIVNKDDEKRN
 NIESVVCNIVNKTVKVGVVPADQWLRVRAEYLQNRNTNETHQSEKQSTQSQQIDIAQKA
 KDLFGEETVHLVDED*

Sequence 1305

Contig_0584_pos_1105_725,
 45 is similar to (with p-value 3.0e-52)
 >pir:pir|S13788|S13788 recM protein - Bacillus subtilis
 atgcattatccagaacctatatcaaagcttatcgatagttttatgaaactgccaggcatt
 ggaccaaagacggctcaacgtctggcttttcatacttttagatatgaaagaagacgatgtt
 50 gttaaagtttgctaaagcactagttgatgttaaaagagaacttacctattgtagtgtttgt
 gggcatattacagaaaatgatccttggtatatatgtgaagataaacagcgagatcggtct
 gtcatatgtgtagttgaagatgacaaggatgtcatagcaatgaaaaaagtcgtgaatata
 aaggtttatatcacgtgcttcattggttcgatttcaccaatggatggtattgggcctgaag
 acatcaatatacctgcattag
 55

Sequence 1306

MHYPEPISKLIIDFSFMKLPGIGPKTAQRLAFHTLDMKEDDVVKFAKALVDVKRELYCSVC
 GHITENDPCYICEDKQRDRSVICVVEDDKDVIAMKKCVNIKVIYITCFMVRFHQWMVLGLK
 TSIYLLH*

Sequence 1307

Contig_0585_pos_2660_3478,

putative peptide of unknown function

5 atgaaaataaatgttttatgcgagaagagggacaatatggatattaaacaatcttcagag
 aaacaaggtcgaccgcatcatttatcagacagtaggacagttttaaagaatatttata
 ttaataaccagcttatattttattacaagtatcgtaccaatcattgttgttttcgggtca
 ttagggtacactgccatgataacacaacaggcaccaccacaatggttgatcatttttca
 ttaagtttaagttttgtgattgctcaaggtctaataattagttatcttttataaaatgcat
 10 caatctgtaataaatgatgtgatgaagcaacaatggatagttgcaaagaataaaataatt
 aaaattgtaatagttgagttgtcgatatatttattattacttataatgcgggtgattgga
 acatcattacctaatacatttaagttatcatctcacgcaatacgaacaacgtacgtaggg
 ctattttaaatacccatatgtgtttagttacttttatatccatggtattcttacgtcca
 atggtagaacaaatcatttatagatatctcatcatccatgaattaggaaaagtatggaat
 15 agacaatttgtaattggttgtctattcttattgaaacgatcgacatgtttacgacatg
 tcacgcgattttgaaattttccatatatcggtattgtacggcagctacaatactatat
 attaaatcgcgggataatttaattgtcgcttatataattcaagtgattttgcagtgatc
 ctttttatagaattttatgtaagtataccaacttttaa

20 Sequence 1308

MKINVLCERDNDIKQSSEKQGRPHLSDSRTVLKRNFIIPAYILLOSIVPIIVVFGS
 LGITAMITQQAPPQWLYHFSLSLSFVIAQGLILVIFYKMHQSVINDVMKQWIVAKNKII
 KIVIVAIVVYLLLLIMRVIGTSLPNHLSYHLTQYEQRTLGLFKSPYVLLVTFISMVFLRP
 MVEQIIYRYLIIHELKGVWNRQFVIGLSILIIETIVHVYDMSSIFEIIPYIIVATAATILY
 25 IKSRDNLIVAYIFQVILQCILFIEILCKYTNF*

Sequence 1309

Contig_0585_pos_8148_8567,

is similar to (with p-value 2.0e-20)

30 >sp:sp|P36922|EBSC_ENTFA EBSC PROTEIN. >pir:pir|C49939|C4993
 9 ebsC protein - Enterococcus faecalis >gp:gp|L23802|ENEEBSA
 _3 Enterococcus faecalis pore forming, cell wall enzyme, reg
 ulatory, and dehydroquinase homologue proteins (ebsA,ebsB,eb
 sC,and ebsD) genes, complete cds with repeat region. NID: g3
 35 88106.
 atgaatacttacgaagttactgacaagcatcaacatggagaagagattgcacaactcgta
 ggtgctaaaatagaagaagtctttaaaacacttgtactagagaattccaatcatgaacac
 tatgtttttgtcattccagtttaataaacccttagatatgaagaaggcggtcatgttgtt
 aatgaaagaataatgaatttaataatgcctctcgatcaattaaaacaagtaacagggtatgtt
 40 agaggaggatgttcacctatcggtatgaaacattcctttaaaacgacgattgatgcttcc
 gctaaaatttagaaaaagtttatattagcggaggtcaaagaggaatgcaaattatcatt
 catgtgaatgatttaattgacatgacaaaggctcaggtagaatctattacacagaattaa

45 Sequence 1310

MNTYEVTDKHQHGEEIAQLVGAKIEEVFKTLVLENSNHEHYVFPVNETLDMKKAHV
 NEKKLNLMLPLDQLKQVTGYVRGGCSPIGMKHSFKTTIDASAKNLEKVYISGGQGRGMQIII
 HVNDLIDMTKAQVESITQN*

50 Sequence 1311

Contig_0585_pos_8862_0,

is similar to (with p-value 0.0e+00)

>sp:sp|P33166|EFTU_BACSU ELONGATION FACTOR TU (EF-TU) (P-40)
 . >pir:pir|A60663|A60663 translation elongation factor Tu -
 55 Bacillus subtilis >gp:gp|Z99104|BSUB0001_113 Bacillus subtil
 is complete genome (section 1 of 21): from 1 to 213080. NID:
 g2632267. >gp:gp|D64127|D64127_6 Bacillus subtilis genes fo
 r RNA polymerase beta subunit, ribosomal proteins L12 and S7
 , elongation factors G and Tu and ribosomal proteins S10 and

L3, partial and complete cds. NID: g1644218.

atggcaaaaaagaaatttgatcgctcaaaagaacatgccaatattgggtactatcggtcac
ggtgaccatgggtaaaacaactttaacagctgctatcgcaactgtattagctaaaaatggt
gacactggtgcacaatcatagcatatgattgacaacgctccagaagaaaaagaacgtggt
5 attacaatcaatactgcacatatcgaaatacacaactgacaaacgtcactatgctcacgtt
gactgccagagacacgtgactatgttaaaaaacatgatcactggtgcagctcaaatggac
ggcgggtatcttagttgtatctgctgctgacggtccaatgccacaaactcgtgaacacatc
ttattatcacgtaacgttgggtgtaccagcattagttgtattcttaacaaaagttgacatg
gtagacgacgaagaattattagaattagttgaaatggaagttcgtgacttattaagcgaa
10 tatgacttcccaggtgacgatgtacctgtaacgtggttctgcattaaaagcattagaa
ggcgtgctgaatacgaacaaaaaatcttagacttaatgcaagcagttgatgattacatt
ccaactccagaacgtgattctgacaaaccattcatgatgccagttgaggacgtattctca
atcactggtcgtgggtactgttctacaggccgtgttgaaacgtggtcaaatcaaagttggt
gaagaagttgaaatcatcggtatgcacgaacacttctaaaacaactgttactggtgtgaa
15 atgttccgtaaattattagactacgctgaagctggtgacaacatcggtgctttattacgt
ggtgttgacagctgaagacgtacaacgtggtcaagtattagctgctcctggttctattaca
ccacacacaaaaattcaaagctgaagtata

Sequence 1312

20 MAKEKFDRSKEHANIGTIGHVDHGKTTTLTAATVLAKNNGDTVAQSYDMIDNAPEEKERG
ITINTAHIEYQTDKRHYAHVDCPGHADYVKNMITGAAQMDGGILVVSAADGPMPTREHI
LLSRNVGVPALVVFLNKVDMVDDEELLELVEMEVRDLLSEYDFPGDDVPVIAGSALKALE
GDAEYEQKILDLMQAVDDYIPTPERDSKPFMPVEDVFSITGRGT VATGRVERGQIKVG
EEVEIIGMHETSKTTVTGVEMFRKLLDYAEAGDNIGALLRGVAREDVQRGQVLAAPGSIT
25 PHTKFKAEVX

Sequence 1313

Contig_0585_pos_7928_6480,
is similar to (with p-value 0.0e+00)
30 >sp:sp|P32397|HEMG_BACSU PROTOPORPHYRINOGEN OXIDASE (EC 1.3.
3.4) (PPO). >pir:pir|D47045|D47045 coproporphyrinogen III ox
idase, protoporphyrinogen IX oxidase - Bacillus subtilis >gp
:gp|M97208|BACHEMEHY_4 Bacillus subtilis penicillin binding
protein 1A (ponA) gene; uroporphyrinogen decarboxylase (hemE
35) gene; ferrochelatase (hemH) gene complete cds, (hemY) gene
, complete cds; ORFA, complete cds; ORFB 5' end. NID: g14304
1. >gp:gp|Z99109|BSUB0006_90 Bacillus subtilis complete geno
me (section 6 of 21): from 999501 to 1209940. NID: g2633260.
>gp:gp|Y14083|BSY14083_8 Bacillus subtilis chromosomal DNA,
40 region 76-78 degrees: between glyB-aprE. NID: g2226224.
atggcgtattttaataaattttaagcaattctgaaaggaagcgtggaaaagtgtagtaag
aaagtggcaattattggagcgggaatcactggtttatctagcgcatatttcattaaaaaa
caagacccttctattgaagtaactatcttcgaagcctcaaatagagtaggtggaaagatt
caaacatatagatcagatggttacacaattgagttaggccctgagtccttatttaggtcgt
45 aagacaattatgactgatgtggcaaaagatattggattagaaaatgaccttatacaaat
actactggccaatcttatatttttgctaaaaataaattatatcctattcctggtggctca
attatgggaattcctacagatattaaaccatttattaaaaacaagactcatttcacctatt
ggtaaatgaagagcgggcttgatttgtttaaaaaacgatagaaattgaagatgatatt
tctgttggtagtttcttttagacaacgattaggtaatgaagtattagagaacttaattgaa
50 ccactaatgggtggtatttatggcactgatattgatcaattgagcttaatgagtacattt
cctaactttaaggaaaaagaggaacaatttggtagtttgattaaaggaatgaaagacgaa
aaagaacaacgtatttaagaaacgtcaattatatccaggtgctcctaaaggacaattcaaa
cagtttagacacggattgagttcttttatagaggctcttgtaaagatattgaaagtaaa
gggtgccacatagatataacacgccagtcacaaagatatattgatttcgcaaaaagattat
gaaattttattagaagatgacagtaaaagagaaatttaattggcttacttgaacaacacca
55 catcaagtatttctgaactggtttagtcacgatccagcatttgattactttaaaaacatg
gattctactactgtcgaacagttgttttgccctttgatgagaaaaatattaccaatagc
tacgatggaactggctttgttattgcaagaacaagtcaaacggatattactgcatgtact
tggacatcaaagaaatggccatttactactccagaaggtaaaagttttaattcgagcatat

ataggaaaaaccaggtgatactgtagtagatgatcacaccgatgaagaaatagtatcaatt
 gttagaaaagacttaagccaaatgatgaccatctcaggaatcctgattttacaattgta
 aatcggttacctaagagtatgccccaataccacgtgggtcatattaaaatgattaaagaa
 attcaacaacatattaaaacaacttatcctagattacgtgttacaggggcaccggttgaa
 5 gctgtcggtttaccagactgcatacaacaaggtagaatgcagttgatgaaatattagaa
 gagttataa

Sequence 1314

MAYFNKFLSNSERKRGKVSKKVAIIAGITGLSSAYFIKKQDPSIEVTIFEASNRVGGKI
 10 QTYRSDGYTIELGPESYLGRKTIMTDVAKDIGLENDLITNTTGQSYIFAKNKLYPIPGGS
 IMGIPTDIKPFIKTRLISPIGKLRAGLDLFKKPIEIEDDISVGSFFRQLRGNEVLENLIE
 PLMGGIYGTIDIDQLSLMSTFPNFKEKEEQFGSLIKGMKDEKEQRIKKRQLYPGAPKGQFK
 QFRHGLSSFIEALVKDIESKGVHIRYNTPVKDILISQKDYEILLEDSDSKEKFNGLLVTTTP
 HQVFLNWFSDHPAFDYFKNMDSTTVATVVLAFDEKNITNTYDGTGFVIARTSQTDITACT
 15 WTSKKWPFTTPEGKVLIRAYIGKPGDTPVDDHTDEEIVSIVRKDLSQMMTISGNPDFTIV
 NRLPKSMPQYHVGHKMIKEIQQHIKTTYPRLRVTGAPFEAVGLPDCIQQGKNAVDEILE
 EL*

Sequence 1315

20 Contig_0585_pos_6344_5928,
 putative peptide of unknown function
 atgacagacattataattgtacactcaaaacatggtaattctaaaaatcattgggtatgaa
 tgggttaaggcataatttaacttttgggaagggtatgatgtttctttattcaatcttgaagca
 aatgatcatgctcaaatgtatgagtggttaaatgaaatgaaacaacaactacatatccgt
 25 aaaaaagatacatattttgtgacccacggatttggctcaatcgctgctttaaaatttctt
 gcagaaacgcacatcacattgaaggtttcttttagtatcgaggatttaagaagatgca
 caagatatagacgaagatgtagatttaaaagggttaaccatcgattacgataaaataaaa
 gagcaagtagataaattttatggactcacgtctaaagatgatcaatatgtttcataa

Sequence 1316

30 MTDIIIVHSHKGNKSNHWEWLRHNLTLLEGYDVSLFNLEANDHAQIDWVNEMKQQLHIR
 KKDTYFVTHGFGSIAALKFLAETHHHIEGFFSIAGFKEDAQDIDEDVDLKGVTIDYDKIK
 EQVDKIFYGLTSKDDQYVS*

Sequence 1317

35 Contig_0585_pos_2551_974,
 is similar to (with p-value 0.0e+00)
 >gp:gp|Z99107|BSUB0004_107 Bacillus subtilis complete genome
 (section 4 of 21): from 600701 to 813890. NID: g2632866. >g
 40 p:gp|Y15254|BSYERABCD_4 Bacillus subtilis 13kB DNA fragment,
 from yerA to sapB gene. NID: g2577959.
 atgtcagttttaactgtcatgcaattcatagtcattatcatcatgattgtgttatta
 acgattatgattcttggggttatttgggtatttaaaagacaaagggtcaaaatcaacacagt
 gtactaagaaattttcctgttttgggtcgaatacgttatatttctgaaaaaatcggtccc
 45 gaattaagacaattttcttcgctaacgataatgaaggtaaacctttttcacgaagtgat
 tataaaaaatattgttttagctggaaaaataaaatcaagaatgactagtttcggtacaggt
 aaggattatgaagaagggttttatattcaaaatcagatgttccacttcaagcaactgaa
 ttacatatcgatcactactgaattcatttctacatttttatcatattgagaatgagcgc
 ctatttagtagagaagaatacagaaaaagcgctcaggttgatccgtttttcttaactgat
 50 gaacatgcagtagtattgggctctaaccttaagcatccctttaaatacaaacgcttagtt
 ggtcaatctgggatgagttatggcgctttaggtaaaaatgcaattactgcactgtcaatg
 gggtagctaaagctggtacatggatgaatacaggtgaaggtggattatctgaatatcat
 ttgaaaggtaattggtgacatcatctatcaaatgggtccaggactctttggggaagagat
 catgattggcaatttttaataagagacatgtttatcaatcttgcgaacacaataatgtacgc
 55 gcatttgaaattaagtttagctcaagggtgctaaaacacgtgggtggacatatggagggaaac
 aaagtcacagaagagattgcacgcattagaaatgtgaaaccatatgaaactattaattca
 cctaactcgttttgattttattaaaaatccaacagatttactgaatttcgttaatcattta
 caatcgataggtcaaaaacctgtcggcttcaaaattgtgtcagtaagttgaagaaata
 gaggcgtagtttaaaccaatggtagagatagacacctatccaagctttattactgttgat

ggtggtgaaggtggtacaggcgctaccttccaagagcttgaagatggtggtggtttaccg
 ttattttacagcacttcctatcgtttcaagtatggtagaaaagtatggcataagaaacaag
 gttaaaaattttgcgtccggttaaattagtgactccagataaaaatcgcaattgcattagga
 ttaggtgcggatctcgtaaatattgctagaggtatgatgataagtgtaggatgcatcatg
 5 agtcaacaatgtcattttaaatacatgtccagttggagtagcaacaaccgatcctaaaaaa
 gaaaagggaacttattggtgatgaaaaacaataaccgtgttacaattatggtacaagtttg
 catgaaggtttatttaacatcgctgcagctgtaggtgttcatagtccaacggagattact
 tccgaccatattatctatagacaattagatggcactacaacgtccattcaggattataaa
 cttaaattaatttcttaa

10

Sequence 1318

MSVLTVMQFIVNIIIMIVLLTIMILGVIWLFKDKGQNHSLRNFPVLGRIRYISEKIGP
 ELRQYFFANDNEGKPFSSRDYKNIVLAGKYKSRMTSFGTGKDYEEGFYIQNTMFPLQATE
 LHIDHTEFISTFLYHIENERLFSREEYRKSQVDPFFLTDEHAVVLGSNLKHPFKIKRLV
 15 GQSGMSYGALGKNAITALSMGLAKAGTWMNTGEGGLSEYHLKNGDIIYQIGPLFGVRD
 HDGNFNRFMFINLAEHNNVRAFEIKLAQGAKTGGHMEGNKVTEEIARIRNVKPYETINS
 PNRFDFIKNPTDLLNFVNLHLSIGQKPVGFKIVVSKVEEIEALVKTMVEIDTYPSTFIVD
 GGEGGTGATFQLELDGVLPLFTALPIVSSMLEKYGIRNKVIFASGKLVTPDKIAIALG
 LGADLVNIARGMMISVGCIMSQQCHLNTCPVGATTPDKKEKGLIVDEKQYRVNTYVTSL
 20 HEGLFNIAAAGVHSPTEITSDHIIYRQLDGTTSIQDYKLLIS*

Sequence 1319

Contig_0586_pos_4250_3699,
 is similar to (with p-value 2.0e-42)
 25 >sp:sp|P44463|LIPA_HAEIN LIPOIC ACID SYNTHETASE (LIP-SYN). >
 pir:pir|G64043|G64043 lipoate biosynthesis protein A (lipA)
 homolog - Haemophilus influenzae (strain Rd KW20) >gp:gp|U32
 688|U32688_5 Haemophilus influenzae Rd section 3 of 163 of t
 he complete genome. NID: g1572966.
 30 gtgtatgcagaaacagtagctaaagtaagagaaagaaatccatttacaacaatagaaatt
 ttaccatctgacatgggtggcgattatgaagcccttgaacattaatggcttctagacca
 gacattcttaatacacaacattgaaacggttcgctcgcttaacaccaagagttcgagctcga
 gcaacttacgatagaactttacaatttttacgtcggttctaaagaattacaacctgatatt
 ccaacaaaaatcaagtttgatggttgggttaggtgaaacgatggaagaaatttatgaaacg
 35 atggatgattttacgcgctaataatgatgttgatcttaactataggtcaatatttacaaccg
 tctcgaaaacatttgaaagttgagaaatattatacgccattagaatttggtaaaaatgaga
 aagattgcaatggaaaaaggatttaaacattgtcaagcaggaccttttagtaagaagctca
 tatcatgctgatgagcaagtgaatgaagcagctaaagagaaacaacgccaaggtgaagaa
 caactcaattaa

40

Sequence 1320

VYAETVRKVRERNPFTTIEILPSDMGGDYEALETLMASRPDIILNHNIEVRRLLTPRVRRAR
 ATYDRTLQFLRRSKELQPDIPTKSSLMVGLGETMEEIYETMDDLRLANDVDILTIGQYLQP
 SRKHLKVEKYTTPLEFGKMRKIAMEKGFKHCQAGPLVRSSYHADEQVNEAAKEKQKQGE
 45 QLN*

Sequence 1321

Contig_0586_pos_3512_3123,
 putative peptide of unknown function
 50 atgattaaagtcgaccaacaatattttgaattgatagaagaatataagagaatgttttgat
 gaggaatatattttagctaggtatttcggatatattagacaaatatgattatgtcgttaggt
 gactatggttacgatcaattacgcttaaaaggattttataaagatagtaataaaaaaggca
 gaaataagtaaacgattttcaagtatacaagattatatactagaatattgtaattttggt
 tgtccttattttgtagtcagacgattgtcaccacaaatgaatttattgaagaaatagatgat
 55 aaagaagatatcattgataaattacatgatgttaagattcaacctactattcaagacaca
 gaaaaacatacccaagctatagatcaatag

Sequence 1322

MIKVDQQYFELIEEYRECFDEEIFSARYSDILDKYDYVVGQYGYDQLRLKGFYKDSNKKK

EISKRFSSIQDYILEYCNFGCPYFVVRRLSPNEFIEEIDDKEDIIDKLHDVKIQPTIQDT
EKHTQAIDQ*

Sequence 1323

- 5 Contig_0586_pos_1348_452,
is similar to (with p-value 3.0e-63)
>gp:gp|D86240|D86240_1 Staphylococcus aureus gene for unknown
function and dlt operon dltA, dltB, dltC and dltD genes, com
plete cds. NID: g1405333.
- 10 atgtgggaacacgatttaacgcccattgtctagagaatcattttcttgctaacggtgaagat
gcaacagcattgtgtgacacattgagtgagcatattgatgaagaagtatttctcagggca
caacaacttaagtgtattgctaataatggcagtaggttttgacaattatgatatttcatta
gcaaagaacatgggtgtagttgttaccatacacctcacgtactgacagagacaactgcc
gaactagggtttacattgatgcttactgtagctcgtagaatcattgaagcgacatcatat
15 attcaagagggttaagtggaaaagttggggaccctacttattatcaggaaaagatgtatac
ggtgcaactgttgggtatttttggaaatgggtgatataggaaaagcgtttgcacgtcgctta
caaggggtttgatcacgataatataatcaccaatcgcaaacgtgacttaaatgctgaaaga
gatttaaatgctacatatgttaacgttttaaatctttacttgaacaaagtgttttattatt
tgcacagcacctttaactaaggaaaactgagaatcaatttgatgctcgagcttttaataaa
20 atgaaaaatgatgctgtcttcattaatattggaagaggtgaaattgtagatgaagaagca
cttttagaagcattaaaaaatcatgagatacaagcctgtggttagatgttacgcgtcaa
gaacctattcaacctaatcatccaatactgaaattacctaacgctgtggtgttacctcac
ataggaagtgcattcccaagtcactagaaatcgaaatggtacaactttgtatagataatatt
aaagcagatttaataatgatgcaccaataacccaataaacctctttacacttttaa
- 25

Sequence 1324

- MWEHDLTPMSRESFLANVEDATACVITLSEHIDEVFLRAQQLKVIANMAVGFDNIDISL
AKKHGVVVTNTPHVLTTETTAELGFTLMLTVARRIIIEATSYIQEGKWKSWGPYLLSGKDVI
GATVGIFGMGDIGKAFARRLQGFDAIIYHNRRKDLNAERDLNATYVTFKSLLEQSDFI
30 CTAPLTKETENQFDARAFNKMKNDAVFINIGRGEIVDEEALLEALKNHEIQACGLDVTRQ
EPIQPNHPILKLPNAVVLPHIGSASQVTRNRMVQLCIDNIKAVLNNDAPITPITSLHF*

Sequence 1325

- 35 Contig_0589_pos_1181_1807,
putative peptide of unknown function
- atgaattttaaaaagactgtagcaattgtcctaacgctcagcagtggttattagctggatgt
actatagataaaaaaagaaattaaaaaatatgatgatcaagtacaaaagctatggaccaa
gagaaaaaccggttaatacagtaagtaaaaaataaacgaattagaagagaaaaagcaaaaa
ttattttaaaaaggtaaatgataaagatcaaaagcacacgtaaaaaagcagctgaagatata
40 gttgaaaaatgtaaaacaaagacaaaaagaatttgaaaaagaagagaaggctctagataat
tctgaaaaagcattttaacaagccaagcaatatcttgaacatgtagaaaacaaagcaaaag
aaaaaagaagttgaacaacttgatagtgctattaaagaaaaatataaatcacatgatgct
tatgcaaaaggcttacaaaaaagcacttaataaggaaaaagaactgttttcttatttgat
gaagataatgcaacacaatcggaagtagacggaataatcgaaagatctttctaaagcatat
45 aaagaaatgaataataaatttaattgcttactcaaaagccattgagaaagtaaaaaagagaa
aaacaagatgtagaccaattaaaaataa

Sequence 1326

- 50 MNFKKTVAIVLTSVLLAGCTIDKKEIKKYDDQVQKAMDQEKTVNQVSKKINELEEKQK
LFKKVNDKQDSTRKKAEDIVENVKQRQKEFEKEEKALDNSEKAFKQAKQYLEHVENKAK
KKEVEQLDSAIKEKYKSHDAYAKAYKKALNKEKELFSYLNEDNATQSEVDGKSKDLSKAY
KEMNNKFNAYSKAIEKVREKQDQVQLK*

Sequence 1327

- 55 Contig_0589_pos_1978_3090,
is similar to (with p-value 0.0e+00)
>pir:pir|S10798|DEBSPF pyruvate dehydrogenase (lipoamide) (E
C 1.2.4.1) alpha chain - Bacillus stearothermophilus >gp:gp|
X53560|BSPDMC_3 B. stearothermophilus pdhA, pdhB, pdhC, pdhD

genes for pyruvate dehydrogenase multienzyme complex (E.C. numbers 1.2.4.1, 2.3.1.12, 1.8.1.4). NID: g40038.

```

atggctcctaagttacaagcccaattcgatgcagttaaagttttaaatgagactcaatcg
aaatttgaaatgggttcaaattttggatgaagacggaatgtcgttaatgaagacttagta
5 cctgatttaacagacgaacaatttagtgaattaatggaaagaatggtatggactagaatt
cttgatcaacggttctatttctgtaaatagacaaggacgttttaggtttctatgcaccaaca
gcaggacaagaagcttcacaatttagcatctcagtatgctttagaagtgaaagacttcatt
ttacctggttatcgtgatgtgcctcagattatttggcatggcttacctcttacagacgca
10 ttcttattctcaagaggacacttcaaaggtaaccaattccctgaggaggttaatgcactt
agccctcaaattattatcggtgcacaatatattcaaactgccggtgtagcggttggtactt
aaaaaacggtggcaaaaatgcagtcgcaattacttatacaggtgatgggtggttcacacaa
ggtgacttctatgaaggaattaactttgcatctgcatacaaaagcacctgcaatttttgta
attcaaaacaataactatgccatctctacaccacgtagtaaaacaaacagctgcagaaaca
15 ttagcacaaaaggctattttagtgggtatccctggaattcaagttgatggtatggatgct
ttagctgtttatcaagcaacattagaagcacgtgaacgtgcagtagcaggagaaggtcct
actgttatcgaaactttaacttatcggtatggaccacatactatggctggtgatgacct
actcggtatagaacttcagatgaagatgctgaatgggagaaaaaagaccatttagtacgt
ttcagaaaaatatttagaagctaaaggctcttggatgaagacaaagaaaatgaagtgggt
20 gaacgtgcaaaatctgaataaaagcagctattaaagggtgacaatacagaaaaacaa
actgttacttctctaattggatatcatgtatgaagaaatgcctcaaaatttagcagaacaa
tatgaaatttacaagagaaggagtcgaagtaa

```

Sequence 1328

```

MAPKLQAFDAVKVLNETQSKFEMVQILDEDDGNVVNEDLVPDLTDEQLVELMERMVWTRI
25 LDQRSISLNRQRLGFYAPTAGQEASQLASQYALESEDFILPGYRDVPQIIWHGLPLTDA
FLFSRGHFKNQFPPEGVNALSPOIIIGAQYIQTAGVAFGLKKRGKNAVAITYTGDGGSSQ
GDFYEGINFASAYKAPAFIVIQNNNYAISTPRSKQTAAETLAQKAISVGIPGIQVDGMDA
LAVYQATLEARERAVAGEGPTVIETLTYRYGPHTMAGDDPTRYRTSDEDAEWKDKPLVR
30 FRKYLEAKGLWNEDKENEVVERAKSEIKAAIKEADNTEKQTVTSLMDIMYEEMPQNLAEQ
YEIYKEKESK*

```

Sequence 1329

Contig_0589_pos_3094_4071,
is similar to (with p-value 0.0e+00)

```

35 >pir:pir|C36718|C36718 pyruvate dehydrogenase (lipoamide) (E
C 1.2.4.1) E1 beta chain precursor - Bacillus subtilis >gp:g
p|AF012285|AF012285_34 Bacillus subtilis mobA-nprE gene regi
on. NID: g3282109. >gp:gp|M57435|BACPYDHY_3 B.subtilis pyruv
ate dehydrogenase complex genes, complete cds; PAL-related l
40 ipoprotein (slp) gene, complete cds, lysine decarboxylase (c
ad) gene, partial cds. NID: g143375. >gp:gp|Z99111|BSUB0008_
131 Bacillus subtilis complete genome (section 8 of 21): fro
m 1394791 to 1603020. NID: g2633699.
atggcacaaatgacaatggttcaagcgattaacgatgcgcttaaaagtgaactcaaaaga
45 gacgaagacggttttagtttctggtgaagacggttggtgtaacggtggtgtattccgtggt
actgaagggtttacaaaaagaatttggcgaagatcgagatttgcatacaccatttagcagag
tctggaattggtgggcttgcaactaggcttagcagtgactggcttccgtcctgttatggaa
attcaattcttaggattcggttatgaagtatttgacgaagtagctggtcaaattgctcgt
actcggttccgttcagggtggaactaaaccagcgctgttacaattcgtagacaccttttggg
50 ggtggcgtccacactccagagttgcatgctgataatttagaaggtagcttagctcaatca
cctggtttgaaagtagttattccatcaggtccttatgatgctaaaggattattaatttct
tctattcaaagtaatgatccagttgtatatctagaacatatgaaatttatcggtctttc
cgtgaagaggttcctgaagaagaatacaaaaattgacattggaaaagccaatgttaaaaaa
gaaggtaatgatattactctaatactcttacggggcaatggtacaagaatcactaaaagct
55 gctgaagagttagaaaaagatggttattcagttgaagttattgacttacgtactgtacaa
ccaattgatatagatacttttagtagcatcagttgagaaaactggacgtgctgtagttgta
caagaagcacacgtcaagctggtgtgggtgcacaagtggcagcagaattagcagagcga
gcaattctttcattagaagctccaatagctcgagtagccgcatcagatacaatttatcca
ttactcaagctgaaaacggttggttaccaaataaaaaagatattatagagcaagctaag

```

gcaactttagaattctaa

Sequence 1330

5 MAQMTMVQAINDAKSELKRDEDVLVFGEDVGVNGGVFRVTEGLQKEFGEDRVFDTPLAE
SGIGGLALGLAVTGFRPVMEIQFLGFVYEVFDEVAGQIARTFRSGGTPAPVTIRTFFG
GGVHTPELHADNLEGILAQSPGLKVVIPSGPYDAKGLLISSIQSNDPVVYLEHMKLYRSF
REEVPEEEYKIDIGKANVKEGNDITLISYGAMVQESLKAAEELEKDGYSVEVIDLRTVQ
PIDIDTLVASVEKTGRAVVVQEAQRQAGVGAQVAAELAERAILSLEAPIARVAASDTIYP
10 FTQAENVWLPNKKDIEQAKATLEF*

Sequence 1331

Contig_0589_pos_4202_5503,
is similar to (with p-value 0.0e+00)
>sp:sp|Q59821|ODP2_STAAU DIHYDROLIPOAMIDE ACETYLTRANSFERASE
15 COMPONENT (E2) OF PYRUVATE DEHYDROGENASE COMPLEX (EC 2.3.1.1
2). >pir:pir|S19722|S19722 dihydrolipoamide S-acetyltransfer
ase (EC 2.3.1.12) chain E2 - Staphylococcus aureus >gp:gp|X5
8434|SAPDHDNA_2 S.aureus pdhB, pdhC and pdhD genes for pyruv
ate decarboxylase, dihydrolipoamide acetyltransferase and di
20 hydrolipoamide dehydrogenase. NID: g48871.
gtggcatttgaatttagattacccgatatcggggaaggtatccacgaaggtgaaattggt
aaatggtttattaaagccggcgatacaattgaagaagatgatgtattagcagaagttcaa
aatgataaatctgtagtagaaattccttctccagtaagtggtactgttgaagaagtgta
gtagatgaaggaacagtggtcagtagtaggagatgtcatcggttaaaattgatgcacctgat
25 gcagaagaaatgcaatttaaaggtcatggcgatgatgaggattctaagaaagaagaaaa
gaacaagaatcaccagtgcaagaagaagcttcacactcaatcacagaagaaagacagaa
gtagatgaaagtaaaactgttaaagcgatgccgtcagtgcgtaagtatgcacgtgaaat
gggtgtcaatattaaagctgtaaatgggtcttggtaaaaatggacgaatcacaaaagaagac
atcgatgcatacttaaattgggtgtagttccgaagaaggttcaaactagcgcagcatct
30 gaatcaacttctagtgatgtcgttaattgcttctgcaacacagcattaccagaaggcgac
ttccctgaaactacagaaaaataacctgcaatgcgcaaagcaattgctaaagcaatggtt
aattctaaacacactgcacctcatgttacattaatggatgaaattgatgtgcaagaatta
tgggatcaccgtaagaaatttaaagaattgctgctgaacaaggtacaaaacttactttc
ttaccatattgtttaaagcattagtttctgcacttaaaaaatatccagcacttaatact
35 tctttcaatgaagaagctggagaggttgtagacacaaacattactggaatattggtattgct
gcagatacggataaaggattattagtagcagtagttaaacatgccgatcgtaaatcaata
ttcgaaatttctgatgaaattaatgaactagctgtaaaagcacgtgatggttaaattaact
tcagaagaaatgaaaggtgcaacatgcacaattagtaatatcggttccgctggtggacaa
tggttctactccagttatcaatcaccagaagtagctatcttaggaattggccgatcgct
40 caaaaacctagcttaaagatggagaaattgtagctgcaccagtgtagctttatcatta
agctttgaccatagacaaatcgatggtgctactggacaaaatgctatgaatcacattaaa
cgcttattaaataatccagaattattattaatggaggggtaa

Sequence 1332

45 VAFEFRLPDIGEGIHGEIVKWFIKAGDTIEEDDLAEVQNDKSVVEIPSPVSGTVEEVL
VDEGTVAVVGDVIVKIDAPDAEEMQFKGHGDDDESKKEEKEQESPVQEEASSTQSQEKTE
VDESKTVKAMPSVRKYARENGVNIKAVNGSGKNGRITKEDIDAYLNGGSSEEGSNTSAAS
ESTSSDVVNASATQALPEGDFPETTEKIPAMRKAIKAMVNSKHTAPHVTLMDVIDVQEL
WDHRKKFKEIAAEQGTKLFLPYVVKALVSALKKYPALNTSFNEEAGEVVHKHYWNIGIA
50 ADTDKGLLPVVKHADRKSIFEISDEINELAVKARDGKLTSEEMKGATCTISNIGSAGGQ
WFTPVINHPEVAILGIGRIAQKPIVKDGEIVAAPVLALSLSFDHRQIDGATGQNAMNHIK
RLNPNPELLMEG*

Sequence 1333

55 Contig_0589_pos_5508_0,
is similar to (with p-value 0.0e+00)
>pir:pir|S19723|S19723 dihydrolipoamide dehydrogenase (EC 1.
8.1.4) - Staphylococcus aureus >gp:gp|X58434|SAPDHDNA_3 S.au
reus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, d

ihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase. NID: g48871.

atggtagttggagatttcccaattgaaacagatactattgtaataggagcaggtccaggt
 ggatatgtcgcagccattcgcgcggctcaattaggacaaaaggtaacaatcgttgagaaa
 5 ggtaatttaggtggtgtatgcttaacggttggtgtataccttcaaaagcattactacat
 gcttctcatcgctttgtgaagcgcaaaattcagaaaacttaggggtaattgctgaaagc
 gtttcgttaaaactatcaaaaaggttcaagaattcaagacttctgtagttaataaattaact
 ggcggtgttgaggacttttaaaaggtaacaaagtagagattgttagaggtgaagccttat
 ttcgttgataacaatagtttacgtgtcatggacgaaaagagtgcctcaacttacaatttc
 10 aaacatgcgattatagctacaggttcaagaccaattgaaattccaaattttgaatttgg
 aaacgtgttatcgattcaacaggagctttaaatctacaagaagtacctaacaactagtt
 gtagttggtggcgatataatcggttctgaattaggtactgcttttgcaactttggtct
 gaagttactatccttgaaggtgcaaaagatattttaggcggatttgaaaagcaaatgaca
 caacctgttaaaaaagggtatgaaagaaaaagggtatcgaaatcgttactgaagcaatggca
 15 aaatctgcagaagaaactgaaaatggtgtcaaaagtaacttatgaggcaaaagggtgaggaa
 caaactatcgaagctgattatgtattagttacagttggcctcgccctaataactgatgaa
 ttaggattagaagaacttggctctgaaatttgctgatcggtgattactagaagtgagacaaa
 caaagtcgtacttctattgaaaatatctttgcgattggagatatgtacctggattacca
 ttagctcacaagctagttatgaaggtaaaagttgctgctgaagcgatagatggtcaagcc
 20 gcagaggtagactatattggtatgccagcagtttgccttacagaaccagaattagcaca
 gtttggttatactgaagctcaagcaaaagaaggtttatcaattaaagcttctaaattc
 ccttatgcagctaattggacgagctttatcattagatgatacaaatggttttgttaagtta
 attaccttaagaagatgatacgttatttggagcacaagttgttaggtactggcgcatct
 gatattat

Sequence 1334

MVVGDFPIETDTIVIGAGPGGYVAAIRAAQLGQKVITVEKGNLGGVCLNVGCIPSKALLH
 ASHRFVEAQNSENLGVIASVSLNYQKVQEFKTSVVKLTGGVEGLLKGNKVEIVRGEAY
 FVDNNSLRVMDEKSAQTYNFKHAIATGSRPIEIPNFEFGKRVIDSTGALNLQEVPNKLV
 30 VVGGGYIGSELGTAFANFGSEVTILEGAKDILGGFEKQMTQPVKGMKEKGIEIVTEAMA
 KSAEETENGVKVTYEAKGEEQTEADYVLTVGRRPNTDELGLEELGLKFADRGLLEVVDK
 QSRTSIENIFAIGDIVPGLPLAHKASYEGKVAAEAIDGQAAEVDYIGMPAVCFTEPELAQ
 VGYTEAQAKEEGLSIKASKFPYAANGRALSLDDTNGFVKLITLKEDDTLIGAQVVGTS
 DIX

Sequence 1335

Contig_0591_pos_416_943,
 is similar to (with p-value 6.0e-32)
 >sp:sp|Q06752|SYC_BACSU CYSTEINYL-TRNA SYNTHETASE (EC 6.1.1.
 16) (CYSTEINE--TRNA LIGASE) (CYSRS). >pir:pir|C53402|C53402
 40 cysteine--trna ligase (EC 6.1.1.16) - Bacillus subtilis >gp:
 gp|D26185|BAC180K_156 B. subtilis DNA, 180 kilobase region o
 f replication origin. NID: g467326. >gp:gp|L14580|BACGLUSYN_
 6 Bacillus subtilis glutamyl-tRNA transferase (gltX), serine
 45 acetyltransferase (cysE), and cysteinyl-tRNA synthetase (cy
 sS) genes, complete cds's. NID: g289278. >gp:gp|X73989|BSCTS
 _1 B.subtilis gene for cysteinyl-tRNA synthetase. NID: g4993
 02. >gp:gp|Z99104|BSUB0001_94 Bacillus subtilis complete gen
 ome (section 1 of 21): from 1 to 213080. NID: g2632267.
 50 atgattagtggtcattatcgtagcccaataaaactacaatttagaattagtaggtgcggcg
 cgaagtggctcttgaacgtatagctaatagctacaagtttaattgaggaaaagagaacaaat
 gcctcagattcttgaagaacaatcagaatatatacaacaataagataaaataactaaatcaa
 tttgaaacggtaattgatgatgactttaatactgctaatacagtaactgcatggtatgac
 tttagctaaacttgcaataaatatgtattagaaaatacaacttcaacaaaagttttaaat
 55 agatttaaagaaggtgtacagcatttttagtgacgtccttgggtgtaccacttaagagtaaa
 gaaactgaagagttactagatgaagacattgaacaattgattgaggagcgtaatgaagca
 cgtaaaaataaagatttcgctcgagcagatgaaattagagatatgttaaaagcacgtcat
 atcattttagaagatacccccaaggtgtaagatttaaacgtggctaa

Sequence 1336

MISVHYRSPINYNLELVGAARSGLERIRNSYKLIEREQIASDLEEQSEYIQQIDKILNQ
FETVMDDDFNTANAVTAWYDLAKLANKYVLENTTSTKVLNRFKEVYSIFSVDLVGVPLKSK
ETEELLEDIEQLIEERNEARKNKDFARADEIRDMLKARHIILEDTPQGVRFKRG*

5

Sequence 1337

Contig_0591_pos_948_1334,

putative peptide of unknown function

atgaacgtaaaacttcttaatcctttaacattggcatatatgggtgatgcagtacttgat
10 caacatgtgcgtgaatatatcgtgctaaaattacaaagtaaaccctcctcgtttgcaccaa
gtatcgaaaagttacgttttcagcgaaaagtcaagctaagacttttagagtatttgtagat
attgactgggttacagaggaagagctaagtgttttaaaacgaggacgtaacgctaaaagt
tatacaaaagctaaaaatactgacattcaaaacttatcgtaaaagttcagcgtagaagct
gttatcggtattttatatttagaccatcaatcagaacgattagaaaacttattagaaca
15 attgtaggatagtggtgaaaggtag

Sequence 1338

MNVKLLNPLTLAYMGDAVLDOHVREYIVLKLQSKPPRLHQVSKSYVSAKSQAKTLEYJLLD
IDWFTEEELSVLKRGRNAKSYTKAKNTDIQTYRKSSALEAVIGFLYLDHQSERLENLLET
20 IVRIVDER*

Sequence 1339

Contig_0591_pos_2685_3260,

putative peptide of unknown function

atgaatctgaataaacagcaacatgaatatacagcactgtgtctatcgcaaacagaaaat
25 aaatcttctgaagaactatttgagtcctttaatagaagagctaaagccactgatttacaat
aaaataaggtatatctccataataagtatgacattgaagacatgtatcaagagattggt
attaaattctaccgtgccttgcaaaaattcgactatcaacaaggtgtaccaatagaacac
tataattttttttaattcggttcggttaaatatgactatcttagaaaagtaaaagcgaat
30 tataaacgctcaacctctacttggttaatgaatacattgtgaatataacgctactttggca
ttaaacgatatagaagatcgataattagaaaagaattaacattagcttttaaaagaagc
gaagtcaaaactcagtcgaatggaaagacgtatcattcgattactacttaatgattacaag
ccaaaggagattgctatggtttttaatttggaatccaaagttggtttataatgcgattcaa
cgtagtaaatgtaaaccttaaaagaagttttgaataa
35

Sequence 1340

MNLNKQQHEYTALCLSQTENKSSEELFESLIEELKPLIYNKIRYISHNKYDIEDMYQEIV
IKFYRALQKFDYQQGVPIEHYIYFLIRSVKYDYLRKVKANYKRQPLLNVNEYIVEYNATLA
LNDIERSIIRKELTLAFKRSEVKLSRMERRIRLLLLNDYKPKEIAMVLNLESKVVYNIAIQ
40 RSKCKLKRSFE*

Sequence 1341

Contig_0591_pos_4275_3676,

is similar to (with p-value 3.0e-28)

45 >gp:gp|L14580|BACGLUSYN_7 Bacillus subtilis glutamyl-tRNA tr
ansferase (gltX), serine acetyltransferase (cysE), and cyste
inyl-tRNA synthetase (cysS) genes, complete cds's. NID: g289
278.

gtgaatgtggaagatatagtgatagtaggttagacacgcagttaaagaagcaattatatca
50 ggtcacgccataaaataagattttgattcaagacggtataaaaaagcaacaaattaacgac
attttaaaaaatgcaaaatcacaaaaattaattgtacaaacggtaccaaatactaaatta
gatttttttagcaaatgcacctcaccagggtgtggctgcttttagtagccccatatgaatat
gcaaaacttcgatgaatttttacaaaaacaaaaagaaaaagccggtattcaactgttatc
atttttagatgggttagaagacccgcataatcttggctctatattaagaacagcagatgct
55 tctggtgttgatgcggttattatacctaaaagacgatcagttgcgctaacacagaccgtt
gcaaaagcttctacaggagcgattcagcatgttccggttataagggttactaatctttcg
aaaactatcgacgaattaaaagacaacggcttttgattgcggggacagaagctataaat
gcaacggattatagactaagaagaatatcactgttgatacaactattattgtatatttaa

Sequence 1342

VNVEDIVIVGRHAVKEAIIISGHAINKILIQDGIKKQQINDILKNAKSQKLIVQTVPKSKL
 DFLANAPHQGVAAALVAPYEYANFDEFLQKQKKKARYSTVIILDGLEDPHNLGSILRTADA
 5 SGVDAVIIPKRRSVALTQTVAKASTGAIQHVPIRVTNLSKTIDELKDNGFWIAGTEANN
 ATDYRLRRISLLIQLLLYI*

Sequence 1343

Contig_0592_pos_10448_10146,
 10 is similar to (with p-value 1.0e-19)
 >gp:gp|AF011545|AF011545_3 Bacillus subtilis SapB (sapB), Op
 uE (opuE), YedA (yedA) genes, complete cds, and YedB (yedB)
 gene, partial cds. NID: g2465554.
 atgactaaagtaaacacgtgaagaagttgaacatattgctaatttagctagacttcaaatt
 15 tctcctgaagaacagaagaatggctaatacttttagaaagtatttttagattttgcgaaa
 caaatgatagtgccgatacagaaggtattgagccaacttatcacgtattagattacaa
 aacgtattacgtgacgataaaagcaatcgaggcattcctcaagaattagcattgaaaaat
 gcgaagaagaactgaagatggccaatttaagtgccatccatcatgaatgggaggacgct
 taa

Sequence 1344

MTKVTTREEVEHIANLARLQISPEETEEMANTLESILDFAKQNDADTEGIEPTYHVLDLQ
 NVLRDDKAIEGIPQELALKNKAKETEDGQFKVPSIMNGEDA*

Sequence 1345

Contig_0592_pos_10144_8687,
 is similar to (with p-value 0.0e+00)
 >gp:gp|AF008553|AF008553_2 Bacillus subtilis Glu-tRNAGln ami
 dotransferase subunits C (gatC), A (gatA) and B (gatB) genes
 30 , complete cds. NID: g2589193.
 atgagtattcggttttgaatctatcgaaaaattaaactgaattaatcaaaaaataaagaatt
 aaaccttctgatgtagtaaaagatatatacgcagctattgaagaactgatccaacaatc
 aagtcattcttagcttttagataaagaaaatgcaataaaaaaagccgaagaattagatgaa
 ttacaagctaaagatcaaatggatggtaaaactatttggaattcctatgggaatcaaatgat
 35 aatatcatcacaagaagatgtagaaactacatgtgcaagtaaaatgttagaaggatttgta
 cctatttatgaatcaactgtaataaactacatgatgaaaacgcggttttaattggt
 aaattaaacatggatgagtttgcaatgggtggctctacagagacttcataattttaagaaa
 acattaaatccttttcgatcacacagcagtagcaggaggatccttcagggtggttctgcagca
 gcggttgacagcaggttttagttccttttagtttagggtcagacactgggtggttctattaga
 40 caacctgcatcttattgtggcgttggttggtatgaaaccaacttatggcgtgtatcacgt
 ttcggtttagttgcattttgcttcttcttagatcaaattggaccaatcacgcgtaatggt
 aaagataacgcatttagtacttgaggcaatttccggtggtgatgcgaatgattctacaagc
 gcacctgttgatgatgtagattttacttctgatattggtaaagatattaaaggtcttaaa
 attgcattacctaagaatatttaggtgaggggtgaagtgaagaagttaagacttctgta
 45 aaagaagcgggttgaaacgttaaaatcacttgggtgctgaagttgacgaagtctcattacca
 aatacaaaaatattggtattccatcatattatgttattgctcatcagaggcttcagcaaat
 tttagcgcgatttgatggtattagatatggatatcattctaaagaagcacaatcgtagaa
 gaattatataaaatgtctagatcggaaggctttgggtgaagaagtcaaaagacgtatcttc
 ttaggtacttttgctttaagctcaggttattacgatgcatactataaaaaatctcaaaaa
 50 gttagaacggttaattaaaaatgattttgacaaagtatttgaatcttatgatgttggtt
 ggaccgacagcactacaacagcatttaattttggcgaagaatttgatgatcctttaaca
 atgtatgcgaatgatcttataactacaccagttaatcttgccggtttacctggtatttca
 gttccttggtgacaatcaaacggacgccaattgggtttacaatttaattggtaaacctttt
 gacgaaaaaacggttatatcggtgctgcttatcaatttgaaacacaatacaacttacatgac
 55 gcatacgaaaatttataa

Sequence 1346

MSIRFESIEKLTELKIKKEIKPSDVVKDIYAAIEETDPTIKSFLALDKENAIKKAEEELDE
 LQAKDQMDGKLFIPMGIKDNIITKDVETTCASKMLEGFVPIYESTVMNKLHDENAVLIG

KLNMDEFAMGGSTETSYFKKTLNPFDHAVPGSSSGSAAVAAGLVPFSLGSDTGGSIR
 QPASVCGVGMKPTYGRVSRFGLVAFASSLDQIGPITRNVKDNLVLEAISGVDANDSTS
 APVDDVDFTSDIGKDIKGLKIALPKEYLGEGVSEEVKTSVKEAVETLKSGLAEVDEVSLP
 NTKYGI PSYYVIASSEASANLARFDGIRYGYHSKEAQSLEELYKMSRSEGFGEVKKRIF
 5 LGTFALSSGYYDAYYKKSQKVRTLIKNDFKVFESYDVVVGPTAPTTAFNIGEEIDDPLT
 MYANDLLTTPVNLAGLPGISVPCGQSNRPIGLQLIGKPFDEKTLYRVAYQFETQYNLHD
 AYENL*

Sequence 1347

10 Contig_0592_pos_8683_7247,
 is similar to (with p-value 0.0e+00)
 >sp:sp|Q45486|YZDD_BACSU PET112-LIKE PROTEIN. >gp:gp|U49790|
 BSU49790_1 Bacillus subtilis PET112-like protein gene, comple
 te cds. NID: gl354210.
 15 gtggaaatcatgcattttgaaacagtaatcggacttgaagttcatggttgagttaaaaacg
 gactcaaaaatggttctcctccatcaccgcacattttggagctgaaccaaattcaaataca
 aatggttatcgacttagcttatccaggtgtattaccagtagttaatagacgtgcagtagat
 tgggcaatgagagcttcaatggcattaaatattgctacaaattcaaaatttgat
 cgtaaaaactatttctatccagataatccaaaagcatatcaaatttctcagtttgatcaa
 20 cctattggagaaaaatggctatattgatattgaagttgatggagaaacaaaacgtatcggt
 attacacgtcttcatatggaagaagatgcaggtaaatacaacacataaagatggttattct
 ctagtagacttaaacggtcaaggtagccattaattgaaattgtatctgaaccgatatt
 cggtcacctaagaagcatatgcttatctagaaaaactacgttcaatcattcaatataca
 ggtgtatctgattgtaaaatggaagaggatccctacgttgtgatgctaataatttcactt
 25 cgtccatattggtcaaaaggaatttggtaaaaaactgaattgaaaaaccttaactcatt
 aactacgttaaaaaaggtttagaatatgaagagaaacgtcaagaagaagaattattaat
 ggtggagagattggtcaagaaacacgtcgatttgatgaatctactggtaaaaacaatttta
 atgcgtgtgaaagaaggttcagatgattatagatatccctgaaccagatattgtacca
 ttatatgtagatgaagattggaaagcacgtgtaagagaaacaattccagaattgccagat
 30 gaacgtaaagctaaatacgtaaatgatcttgactaccagaatatgatgcgcagtatta
 acattaactaaagaaatgtctgatttcttgaaggcgcaattgaccatggtgcagatggt
 aaacttacttccaactgggttaattgggaggtgtaacgagtatcttaataaaaaatcaagtt
 gaattaaaagatacgcaactaacacctgaaaaatttagctggtatgattaaattaatagaa
 gacggaacaatgagtagtaaaatcgctaaaaaagttttccagaactagcagaaaatggt
 35 ggagatgctaacaacaaattatggaagataaaggttttagtacaatttctgatgaagcaaca
 ctacttaatttgaacagatgcattagataataatccacaatcaatagaagattataaa
 aatggtaaaaggtaaagctatgggattcttagtgggccaaattatgaaagcttctaaaggt
 caagctaaccacaaaaaggttaatagcctattaaaaacaagaattagataaaccgttaa

40 Sequence 1348

VEIMHFETVIGLEVHVELKTD SKMFSPSPAHFGAEPNSNTNVIDLAYPGVLPVNNRAVD
 WAMRASMALNMDIATNSKFDRKNYFYPDNPKAYQISQFDQPIGENGYIDIEVDGETKRIG
 ITRLHMEEDAGKSTHKDGYSVLNLRQGTPLIEIVSEPDIRSPKEAYAYLEKLRSIIQYT
 GVSDCKMEEGSLRCDANISLRPYGQKEFGTKTELKLNLSFNYYVKKGLEEYEEKRQEEELN
 45 GGEIGQETRRFDESTGKTILMRVKEGSDDYRYFPEPDIVPLYVDEDWKARVRETIPELPD
 ERKAKYVNDLGLPEYDAHVLTLTKEMSDFFEGAI DHGADVKLTSNWLMMGGVNEYLNKNQV
 ELKDTQLTPENLAGMIKLIEDGTMSSKIAKKVFPPELAENGGDAKQIMEDKGLVQISDEAT
 LLKFVTDALDNNPQSI EDYKNGKGKAMGFLVGQIMKASKGQANPQKVNSLLKQELDNR*

50 Sequence 1349

Contig_0592_pos_7022_6072,
 is similar to (with p-value 2.0e-29)
 >sp:sp|P39074|BMRU_BACSU BMRU PROTEIN. >gp:gp|L25604|BACBMRU
 RBE_1 Bacillus subtilis bmrU, multidrug efflux transporter (bmr)
 55 bmr) and its regulator (bmrR) genes, complete cds, and branched-chain 2-oxo acid dehydrogenase (bfmB) gene, 3' end. NID:
 g2558636. >gp:gp|D84432|BACJH642_251 Bacillus subtilis DNA,
 283 Kb region containing skin element. NID: g2627063. >gp:g
 p|Z99116|BSUB0013_111 Bacillus subtilis complete genome (sec

tion 13 of 21): from 2395261 to 2613730. NID: g2634723.

atgagaaaacgtgcaagaattatatataatccaacatcaggaaaagaactttttaaacgt
gtattaccagatgcactgattaaacttgagaaggcagggttatgaaacgagtgcatatgca
actgaaaaaattggtgatgctacttttgaagctgaaagagcactagaaagtgaatatgat
5 ttactcattgcagctggaggtgacggtacgttaaatgaggtggtcaacggaatcgccgaa
caaccctaactggcctaataatagggtgtaataccaatgggcaccgttaatgactttggaaga
gcacttcatttaccagcgatataatgggggcgattgatgtaatcattgatggtcacaca
accaaggtagatattggaagatgaataatcggtatttcattaacctagctgcagggggg
aaactaacacaagtatcttatgaaacaccaagtaagttgaaatcaattgtaggaccgttc
10 gcgtattacattaaaggattcgaaatgttacctcaaatagaaagcagtagatgtacgtatc
gaatatgatgataacatcttccaaggagaagctttactattccttttaggtttaacgaat
tcaatggctggctttgaaaaattagttccagatgcgaagccttgacgacggttatttcacg
ttaattattgtagaaaaagcaaatcttgctgaattgggtcatattatgacactagcgtct
cgaggtgagcatacaaaacatcctaaagtcatttatgctaaagcgaagtctattaatatt
15 tcatcatttactgatatgcaacttaatgttgatgggtgaatacgggtgggaaattacctgca
aatttccttaatttagaacagcacatagaaatttttacacctaagatgtatttaacgaa
gaactattagaaaatgatacgataactgatattacgcctgataagcaataa

Sequence 1350

20 MRKRARIYNPTSGKELFKRVLPDALIKLEKAGYETSAYATEKIGDATFEAERALESEYD
LLIAAGGDGTLNEVNGIAEQPNRPKLGVIPMGTVNDFGRALHLPDIMGAIDVIDGHT
TKVDIGKMNNRYFINLAAGGKLTQVSYETPSKLKSIVGPFAYYIKGFEMPLQMKAVDVRI
EYDDNIFQGEALLFLGLTNSMAGFEKLVPAKLDDGYFTLIIVEKANLAEIGHIMTLAS
RGEHTKHPKVIYAKAKSINISSFTDMQLNVDGEYGGKLPANFLNLEQHIEIFTPKDVFNE
25 ELLENDTITDITPDQK*

Sequence 1351

Contig_0592_pos_6040_4622,
is similar to (with p-value 0.0e+00)

30 >gp:gp|299108|BSUB0005_71 Bacillus subtilis complete genome
(section 5 of 21): from 802821 to 1011250. NID: g2633055. >g
p:gp|D78509|D78509_9 Bacillus subtilis YfjG-YfjR genes, comp
lete cds. NID: g2780390.
atgccacacatttatattcctaaccctttttgtatgtgggagggaaaaaattgaaacaatt
35 aagaaaaacgaagttaaaacgggaaagggttattgatttaactcatgaggacacggagtt
gttaagttgatcgatatccaatttttattcctaatacttgaattgatgaagaaattaaa
tttaaattaattaaagtgaagaaagaattttgctataggaaaattgatagaggtcataagt
gaaagtgatgatagagtgcaccaccttgtatttattatgcaaagtgtggtggtgtcaa
ttacaacatatagcatatagagcgcaattggatagaaaagagaacaagtagttaatcct
40 tttcatagaaaaggcccttttgagaatacgggtataaaggaaactattggcatgggtcaat
ccctggcgataccgtaataaatctcaaatcctgtagggtcaaagtaactcgaatcaagtt
ataatgggattctatagacaacgttagccatgacattatagatatggatagttgtcttata
caagatagacaacatcaagaagtaataatcgagtgaagtactgggtcaatgaattaaat
atatctatatataacgaaaaaacaaaaacaggtttaatacgtcatttagtagtaaggact
45 gggtatcataccgatgaaatgatgggttatctttgttacaatggagcaacatttaaaca
tcagaactattagtaaaacagctaaaaaaagaatttccaaatataacaagtattaaacag
aatataaacaatagccattcctaattgtataatggggcgctcaatcaatgactttatatggt
aaagataaaaattgaagaccaattaagtgaagtaacttatcatatttctgatttatcatt
taccaaattaaactcatcacaactgaaaaactttatcagcaagctctgaattatgctcaa
50 ttaacaggaaaagaatagtagttggatagctattgtggtataggaacgattggtctatat
atggcaccactagcaaaacatgtttatgggtgtagaagttgttccgcaagccataaaagat
gcggaagacaatgcgactaaaaaccaacttaaaaatacgaactttcgaatgtggaaaagca
gaagtgttatcttaacatggaaatcacaagggttaaaaccaggcgtagtcatggttagat
ccacctagaaaaggatgcgatgaaactttcttaactactcttttaaaattaaatccgaaa
55 aggattgtttatatatcatgtaacccttcaacgcaacaaagagatgcgcaaatattgggt
gaacaatacagagtttagtagaaattacaccagttgatattgtcccaacaacaactcatatt
gagactgtagcattatttgtacgtaaaagacgaagaatga

Sequence 1352

MPHIIFLTFLYVGGKKLETIKKNEVKTKGVIDLTHEGHGVVKVDRIPIFIPNALIDEEIK
FKLIKVKKNFAIGKLIIEVISESDDRVTTPCIYYAKCGGCQLQHMTYRAQLDMKREQVNNL
FHRKGPFFENTVIKETIGMVNPWRYRNKSQIPVGQSNQVIMGFYRQRSHDIIDMSCLI
QDRQHQEVMNRVKYWLNELNISIYNEKTKTGLIRHLVVRTGYHTDEMMVIFVTNGATFKQ
5 SELLVNKLKKEFPNITSIKQNINNSHNVIMGRQSMPLYGKDIEDQLSEVYHISDLSF
YQINSSQTEKLYQQALNYAQLTGKEIVLDTYCGIGTIGLYMAPLAKHVYGVVEVVPQAIKD
AEDNATKNQLKNTTFECGKAEDVILTWSQGIKPGVVMVDPKRGKDEFTLTLLKLNPK
RIVYISCNPSTQORDAQILAEQYELVEITPVDMEFPQTHIETVALFVRKDEE*

10 Sequence 1353
Contig_0592_pos_4484_3936,
putative peptide of unknown function
gtgaatgtgcataaaatagatttatcaggcaacaaatttcaaatccaacgatttgttctg
ttgcaaattgtattggcgctatttacaatactatttacttataaatgggcttatcaaaca
15 acgcatatcattgaacaaaatctgtcatgaatcttatttttgattttaggtttcgca
gtactagtattttgcacgagtttattcatcgtattttgttcattatattttctaaagggt
gaaaaaccatctttaaaatatgataaaaaacaaattattgtacagttctctcagacttgt
tttcacggtggcaatttacaattatcatgatagcaccacttggtatcataagtgcgacc
ttactagcacttattcaaatatattccttctcatctttaatctttatgtttagtatacat
20 acaagttattgtatgatagatgtgttttttagtagcattggcattacaaagcaaattcaaa
tacatacaaacctatggagaagggtttgtatctttatcatcaaagcctactcaaacctat
tatgaataa

Sequence 1354
25 VNVHKIDLSGNKFQIQRFVLLQIVLALFTILFTYKWAYQTTHIEQNLVMNLIFGFVGFA
VLVILHEFIHRILFIIIFSKGEKPSLKYDKNKIIIVQFSQTCFHRWQFTIIMIAPLVIISAT
LLALIQIYSFSSLIIFMSIHTSYCMIDVFLVALALQSKFKYIQTYGEGLYLYHQKPTQTY
YE*

30 Sequence 1355
Contig_0592_pos_3018_1894,
is similar to (with p-value 8.0e-28)
>sp:sp|P23479|SBCD_BACSU EXONUCLEASE SBCD HOMOLOG (FRAGMENT)

35 atgaaaattgtacataccgctgattggcatctgggtaaaattttaaatggaaaacaattg
cttgaagatcaaaaatatattttaactcagtttaaacacatatggagaaagaacagcca
gatttaatatgaattgcaggtgatttgtatgatacctcatatccaagtaaagaagcgata
ggtttacttgaagagactattgaatacctaaatatagaacttaaaattccaataatcatg
ataagcggtaaaccatgatggtagggagagattgaattatggctctaaatggtttgagaat
40 aatcaactttacataagaactcaactagaaaatattgatgatccaatagaattgagtggt
gttcaatttttcactttacctttcgcaactgtgagtgaagtacaaaattattttaaggat
aagcaaatagaacatatcaacaagcattaaacgaatgcttagagcaaatgtctagtcc
atagataataataaggtgaatatattaattggctcatttaactattgaggcggtaaaact
tcagattcggaagaccattaaactattggaacagtagaatcagttgatatgcattctttt
45 cgggtgtttgattatgtaatgctcgggcacctacatcatccatttagtataaaataactct
tttatcaaatatagcgggtcgattttgcaatactctttctctgaagtaaatcaatctaaa
ggatatagagttcttgatattgaaaacaaccaactattaaatgaaaccttcgttccttta
aaacctctaagagaactagaagttattgaaggtgattatgaggatattattcaagaaaga
attaaagtaaaaaataaaaaataattattttcattttaagttaacgaatgtttctcatatt
50 actgatccaatgatgaaactgaaacaaatttatcccaatatattagcactatcgaatgta
gtatttgatcatagtgagaattttagccatgttgaaatcaaaaaacaagatgatcagaca
attatagaaaatttttataaaaaatgatgacagatcaacatctgagtcgaagttcaatcagac
aaaataaagcactgtttaagttttatattggataggagggggtaa

55 Sequence 1356
MKIVHTADWHLGKILNGKQLLEDQKYILTQFKQHMEKEQPDILIVIAGDLYDTSYPSKEAI
GLLEETIEYLNIELKIPIIMISGNHDGRERLNYGSKWFENNQLYIRTQLENIDDPIELSG
VQFFTLPFATVSEVQNYFKDKQIETYYQALNECLEQMSSSIDNNKVNILIGHTIEGGKT
SDSERPLTIGTVESVDMHSFRLFDYVMLGHLHHPFSINNSFIKYSILQYSFSEVNQSK

GYRVLDIENNQLLNETFVPLKPLRELEVIEGDYEDIIQERIKVKNKNNYFHFKLTVNSHI
TDPMMKLKQIYPNIALSNVVDHSENFHSHVEIKKQDDQTIENFYKNMTDQHL SQVQSD
KIKHLLSFILDREG*

5 Sequence 1357
Contig_0592_pos_1869_118,
putative peptide of unknown function
atggagaacttttgcccttttattaaagaaactattgattttgagcaagttgaaactgat
caactccttttaattagtggtataaactggatctggtataaacaatgatttttgatgctata
10 gtatacgcattatacggatggtctcgacaaaactagaaaagaaggagatttaagaagt
cattttgcagacggtaaatcgccaatgtctgtaatttatcaatttaaagttataaatcaa
acttttaaaattcatagagaagcgccatttattaaagaggggaatataaactaaaacacaa
gccaaagttaatatatatgaattagttgataatcaatttgaattaagagaaagttaagt
aatcaaggtaatcaatttatcgtacaattattagcggttaattgctgaacaatttcgtcaa
15 ttatttattttgcctcaaggagaatttataaaagtttcttcagtcataatagtaagacaaa
caatcgattccttagaacactttttaaattagtgagcgatttgatgagattagacatctactt
gtagaaaatgtaaagcaaaaaagtagaaaattgaaaatagatacactcaaatgaaaat
ttatggaaatgatatacatatttaataatgatgaattggccttatataaagaattagag
agttctcagacagataaaaatgattgaaaaattcccacaatttaattgattatggatgcaa
20 attctcaagtcatttgaagaagctaagaataaaaataactaaggaattagatgatttaaat
cataaatataaagtgaatggtgaattaagttagaataactaaaaaattaaaagcggaaaaa
atcaaatgtgacgatttgaaaaagaacaaaatttatattgataaattaaagcaagaatta
aaaatgattcaggaatctaaagtattaatcacttattttactaggttacaaagtttaaaa
aaagataaagatgaattagtgctcattcatgagcaatcaaaattaaacgaacaaaactat
25 cacaatgaaattaaaggttttcaaaaacaactcgaacatttatcaacacgagaaaatgaa
ataactcaatttaattcagtatctagaaaaaaacaaagtttcttcaatcaattagataag
attattagtagttatcaacaaaaaccggttaattgaagaagaaataaaaagattatacagt
gaatataatgattttaataacaaaaaagaagaattgacgaagaagaatgaacaacaagaac
aaagattttgcaattattgaacattacactgaagagatttataagctgaaaaagattata
30 gatgaatctgaaagacaaaaaaaggatgagaaattatttgataaattacaactagataaa
tcactcttatcttagcaaatataaagagaagaagaacagttaaatgaaattgaatcatca
atcacaatatagatgcgactttaattgatttgatgacaaaaaggattttgtaaatgaa
ataaagtcogctatgtcaattggagataacctgtccaatttggtggaatgaaatacattca
ttgggagaacatatgtattttgaatcaattgctcaaaaaataataaaaataaaacggtta
35 gaaagtaagaaggtaaaaattcgtgatgaaataatcaaaatagaaactcgaattgaatct
aacgctttacctccgctaacaacaaacgtcaatgttttttctttatattgttta
actatacgttga

Sequence 1358
40 MENFGPFIKETIDFEQVETDQLFLISGKTGSGKTMIFDAIVYALYGMASKTRKEGDLRS
HFADGKSPMSVIYQFKVNNQTFKIHREAPFIKEGNITKTQAKLNIYELVDNQFELRESKV
NQGNQFIVQLLGVNAEQFRQLFILPQGEFKFLQSNKDKQSILRTLFSERFDEIRHLL
VENVKQEKVQIENRYTQIENLWNDIDTFNDELALYKELESSQTDKMIKFPQFNDYGCK
ILKSFEAAKNKITKELDDLNHKYKVNVELSENTKKLKAEEKIFDDLKKEQNYIDKLKQEL
45 KMIQESKVLITYFTRLQSLKKDKDELVSLEHQSCLNETNYHNEIKGFQKQLEHLSTRENE
ITQFNQYLEKNQVFFNQLDKIISYQQKPVIEEIKRLYSEYNDLITKKEELTKEMNNKN
KDFAIIEHYTEEYKLLKIIDESERQKKDEKLFDKLQDKSSYLSKLKEKKEQLNEIESS
ITNIDATLIDLNDKKDFVNEIKSAMSIGDTCPICGNEIHSLEHIDFESIAQKNNKIKRL
ESKKVKIRDEIIKIETRIESNALPPPKPITSMFFSLYCLTIR*

50 Sequence 1359
Contig_0593_pos_9545_10618,
is similar to (with p-value 0.0e+00)
>sp:sp|P25811|THDF_BACSU POSSIBLE THIOPHENE AND FURAN OXIDAT
55 ION PROTEIN THDF. >pir:pir|JQ1215|JQ1215 hypothetical 50K pr
oteins - Bacillus subtilis >gp:gp|D26185|BAC180K_60 B. subtil
is DNA, 180 kilobase region of replication origin. NID: g467
326. >gp:gp|X62539|BSORIGS_5 B.subtilis genes rpmH, rnpA, 50
kd, gidA and gidB. NID: g40020. >gp:gp|Z99124|BSUB0021_207 B

acillus subtilis complete genome (section 21 of 21): from 39
99281 to 4214814. NID: g2636442.

atgattacaattactttcatatTTTTGagaatagtGCCAAAAGGtaataccaacttctcc
attattatagatttatctcaagcagaagcggttatggattttatcgttccaaaactgat
5 cgagcttctaaggttgcgatgaatcaaatagaaggacgtttaagtgaacttaatacaagaaa
caacgtcaatccatatttagagatactcgcccaagttgaagttaacattgattatccagag
tatgatgatgtagaagacgcaacgacggacttcttactagaacagtctaagcgattaaa
gaagaaatcaatcagttacttgaaacaggagcacaaggtaaaataatgagagaagggtta
tctacagttattgtaggacgtcctaattgttggaagtcttcgatgctaaataaccttatt
10 caagataataaagcaattgtgactgaggtcgctggtacaacaagagacgtgttagaagaa
tatgtcaatgttagaggtgtcccggttacgactttagataactgctgggtatttagggatact
gaagatatcgtagaagattggtgtagaacgttctaggaaagctttaagtgaagcagat
ttaattttattgtgcttaataacaattgaacctctgacggaagatgatcaactttattc
gaagtcatataaaatgaggatgttattgtaataataaaaacagatttagaacagcga
15 ttagatgttagcgaactaagagagatgattggtgatatgccacttatacaacatcgatg
cttaacaagaaggtattgatgaattagaaatacaaataaagatttattcttgggtggc
gaagtacaaaatcaagatatgacttatgtatctaattcacgtcacatttcattgttgaaa
caagcgagacaatcaattcaagatgcatgctgctgagtcctggatcccaatggat
atggtacagattgatttaacacgtacttggaaattctaggagaaattattggagaatca
20 gcgagtgatgaattaatagatcaactatttagtcaatttgtttaggaaaataa

Sequence 1360

MITITFIFLRIVPKGNTNFSIIIDLSQAEAVMDFIRSKTDRASKVAMNQIEGRSLDIKK
QRQSILEILAQVEVNIDYPEYDDVEDATDFLLEQSKRIKEEINQLLETGAQKIMREGL
25 STVIVGRPNVKGSSMLNLIQDNKAIVTEVAGTTRDVLEEYVNVGRVPLRLVDTAGIRD
EDIVEKIGVERSRKALSEADLILFVLNNNEPLTEDDQTLFEVIKNEDVIVIINKTDLEQR
LDVSELREMIGDMPLIQTSMLKQEGIDELEIQIKDLFFGGEVQNQDMTYVSNRHSLLK
QARQSIQDAIDAAESGIPMDMVQIDLTRTWEILGEIIGESASDELIDQLFSQFCLGK*

Sequence 1361

Contig_0593_pos_10691_0,
is similar to (with p-value 0.0e+00)
>sp:sp|P25812|GIDA_BACSU GLUCOSE INHIBITED DIVISION PROTEIN
A. >pir:pir|JQ1216|BWBSGA gidA protein - Bacillus subtilis >
35 gp:gp|D26185|BAC180K_59 B. subtilis DNA, 180 kilobase region
of replication origin. NID: g467326. >gp:gp|X62539|BSORIGS_
6 B. subtilis genes rpmH, rnpA, 50kd, gidA and gidB. NID: g40
020. >gp:gp|Z99124|BSUB0021_206 Bacillus subtilis complete g
enome (section 21 of 21): from 3999281 to 4214814. NID: g263
40 6442.
gtgggttcaagaatatgatgtagtagtcattgggtgctggtcacgccggtattgaagcaggt
ctagcttcagctcgccgggtgctaaaacactgatgttaacaattaatttagataatatt
gctttcatgccatgtaatccatctgtaggtggtcctgcgaaggaatcgttgtagtgaa
atagacgcttttaggtggcaaaatggcaaaaactattgataaaactcacattcaaatgcgt
45 atgcttaatacaggtaaaggtccagctgttagagctttacgtgctcaagcagataaagta
ttatatcaacaagaaatgaagcgtgtacttgagaatgaggataatttagacatcatgcaa
ggatgggttgatgaactcattatagaagataatgaagttaaaggtgttcgtactaatatt
ggtacagaatatcgttctaaagctgtcattattacaacaggtacattcttacgtggagaa
attatactaggaaaacttaaaatattctagtggccctaaccatcaattaccatctgtaact
50 ctacgggataatttaagaaaataggatttgatcgttagatttaaaacgggtacacca
ccacgtgtaaatgcgagaaccatcgattattctaaaactgaaatccaaccaggtgatgat
ataggtcgagcgtttagttttgaaacaaccgaattttatttttagatcaattacctgtgtg
ttaacttatcaaatggagatacacatcaagtcattgatgataacttacatttatctgct
atgtattccggtatgattaaaggtacaggtcctagatatgtccatcaattgaggataaa
55 tttgtccgctttaacgataaaaccaagacatcaacttttcttagaacctgaaggacgtaat
acgaatgaggtata

Sequence 1362

VVQEYDVVVIGAGHAGIEAGLASARRGAKTLMILTINLDNIAFMPCNPSVGGPAKGIVVRE

5 IDALGGQMAKTIDKTHIQMRMLNTGKGPVRLRAQADKVLVYQQEMKRVLENEEDNLDIMQ
GMVDELIIEDNEVKGVRTNIGTEYRSKAVIITGTFLRGEIILGNLYSSGPNHQLPSVT
LADNLRKLGFDIVREKTGTPPRVNARTIDYSKTEIQPGDDIGRAFSFETTEFILDQLPCW
LTYTNGDTHQVIDDNLHLSAMYSGLIKGTGPRYCPSIEDKFVRFNDKPRHQLFLEPEGRN
TNEVX

Sequence 1363

Contig_0593_pos_9014_8610,

putative peptide of unknown function

10 gtgtcagtaactgtaaaaggacaaactgaaacagaatggcttccagtattggatttttaga
aacaaatcttttagcaaaaggttagcgcgacaacatttgatattaataaagctcaaaaacgt
tgtttcgtaaaagctgcagcattacatggcctaggctctttatatatacaacggggaagaa
gttccaagcgctaacgacaatgacattacagaattagaagagcgtatcaaccagtttgta
acttcatctcaagaaaaaggtagagacgcaacgctagacaaaacaatgcgttggttaggt
15 attcaaaacattaacaaagtactaaaaaagatatagcaaatgcacatcaaaaactagat
gcaggactaaaacaattagataaggagaattcaaatgttaaatag

Sequence 1364

20 VSVTVKQGQTEEWLPVLDNRNLSLAKGSATTFDINKAQKRCFVKAAALHGLGLYIYNTEE
VPSANDNDITELEERINQFVTSSQEKGRDATLDKTMRWLGIQNINKVTKKDIANAQKLD
AGLKQLDKENSINVK*

Sequence 1365

Contig_0593_pos_8163_7489,

putative peptide of unknown function

25 atggtagtaataaaaaactacattacagaagatgacgggtacaacaactgtagtcatcaaa
ggagtagaactagataacaaaacatctttacttttagacaacggttacgaagtagaagca
gatgtaagagttgtagatccattcaagattacagataagcagcgtagaaaagtatttgct
ctctgtaacgacatagaagcttacacaggacaaccacgcgactatatgaggtatttgctc
30 atggattacgtagaagttctctatggctatgaaaaacgtctctcattgagtgattgcaca
agagaacaagctaaacaagttatagaagttattcttgactgggtgtttcacacaatatata
ccacttaattataagacaagtgacttactcaaaaatgataaaagcgttcctttactgggtca
acagtcgaatcgtaactgtgtaatatgcggaacgcccacgagcagaacttgcgcatcatcac
acagtaggtcgaggacgtaacagacgaaagatagatcacacagacaacaaagtattagcg
35 ctatgttcaagacatcataaagagcagcaccacaaataggtatagatagttttaatgagaaa
tacaattacatgaaagttgggtgtccgtagatgaacgactcaaccgaatgttgaaagga
gaagtaaatggctga

Sequence 1366

40 MVIKINYITEDGTTTVVIKGVELDNKTSLLLDNGYEVEADVVRVDDPFKITDKQRRKVFA
LCNDIEAYTGQPRDYMYLFDYVEVLYGYEKRLSLSDCTREQAKQVIEVILDWVFHNNI
PLNYKTSDDLKNDKAFLYWSTVNRNCVICGTPRAELAHYHTVGRGRNRRKIDHTDNKVLA
LCSRHHKEHQHIGIDSFNEKYKLHESWVSVDRLNRMLKGEVNG*

Sequence 1367

Contig_0593_pos_7460_6702,

putative peptide of unknown function

50 atgttcgatgatagcaaaatcaagtatatagaagcactgccagaacgagatacaatcatc
actttatgggttaagttgctgacattagctggaaagtataacgaacaaggatcacattatg
ttatccgaaagtctaccctataacgaagaaatgttagctaacgaatttaataagacctatc
aattcaataagattagcgttacaaacattcgaaaagctaagcatgattgaagaagtgaat
gggtgctcttaaggtatctaattgggaaaaacatcagaacatcgaagggttagaaaaagata
agagaacaaaaacggtttgctgtaaacaaaaagcaaaagaaaaaaacaaaaacttttagatagt
cacgtgaagtcacgtgacagtcacgcaacagatatagaagaagataaagaagtagaagaa
55 gaaagagaaaaagaagtagataaagatatcttcaaaaactcaattaattacatcatgagt
aaccttactcataatttaactcctaaccacaaatggaacagataggatatgccattgatgat
attggacaacatgcagatgaagttggtgaagtagctactgattatacaaaagacaaaggt
tgtcatgcaggttacctaatacaagtggttaaaacactgggctaaagagaacgttaagaat
aaaaaagaggctgaaaataaaattaaacctaaaaataaaaaaactgtaacagatgatgta

attgctcaaatggagaaagagctaggagatgaaagttaa

Sequence 1368

MFDDSKIYIEALPERDTIITLWVKLLTLAGKYNEQGYIMLSSESLPYNEEMLANEFNRPI
5 NSIRLALQTFEKLMSIEEVNGVFKVSNWEKHQNIIEGLEKIREQNRLRKQKQRKKQKLLDS
HVKSROSHATDIEEDKEVEEREKEVDKIDFKNSINYIMS NLTHNLTPNQMEQIGYAIDD
IGQHADEVVEVATDYTKDKGCHAGYLIKVLNNWAKENVKNKKEAENKIKPKNKKTVTDDV
IAQMEKELGDES*

10 Sequence 1369

Contig_0593_pos_6696_6343,

putative peptide of unknown function

atgactaaacaacaagccctagaagtaattaagacaattagacatgtatacaacattgac
tttgacagacctaataattagaaacatgggttaacattttgagccaaaatggggattatgaa
15 ccgactaaaaaaacagtaaatgcaatatatcaatgatgctaataccttatccacctagtatt
ccaaacataatgagaaaaagaagtcaaagtcgtaaaagaagagcctgtcgacgaaaaaact
gctagacatcggttgagaatgaaaaatgatccagaatacgtagcacaacgtaaaaagata
ttagacgacttcagaaaagaagttaagtgagtttgagtgagtgacgatgaatga

20 Sequence 1370

MTKQQALEVIKTIRHVYNIDFDRPKLETWVNILSQNGDYEPKTKTVMQYINDANPYPPSI
PNIMRKEVKVVKPEPVDEKTARHRWRMKNDPEYVAQRKKILDDFRKKLSEFGVSDDE*

Sequence 1371

25 Contig_0593_pos_4059_3700,

putative peptide of unknown function

atggaagcaacaaaaatgagagttaaaaataataacttctctattacaccagatgtagta
gagaaaaatgaaagaagcagatatcaatccgatattcttaagacaaagattagcttctggt
tggaagtttgaaagatgcaatagaagcacctattggagtaagacgtagtgagtgaggatagt
30 ttgaaacctaaagaggacgaaattgctagttataaagagagaatggagcaacgcagatta
caagagttgaaacgtaagaaaccacatttattcacagtaaatcaaaaacactctcggtgt
aaatggtgcacgtatcttatggagaatgacatctttcctagaaagggtggttagatcatga

35 Sequence 1372

MEATKMRVKNKYFSITPDVVEKMKEADINPDILRQLASGWKFEDAIEAPIGVRSEWDS
LKPKEDEIASYKERMEQRRQLQELKRKKPHLFTVNQKHSRGKWCTYLMENDIFPRKVRS*

40 Sequence 1373

Contig_0593_pos_3664_3260,

putative peptide of unknown function

atgcatggacttaacggtgtggaagttacagcaaagggttaaaaatgtatatcgtttagtt
cattcaagacgtggtgcggctaaatgggttgctgatgtaaaagcgattgatgggaaaact
45 tggactattgatgataattacgatttttactcattaccagatgaaaatgaagaaaacaaa
aagacggttatatgacaagattaaccacccgctcacattacacatatggagaaatagaagta
attgaattcatagaacagggtcactaaagattataaaccagagttagcatttgcgattggt
aatgcaattaaatatatcagtcgagctaatacgtaagaacggaaaagaagatttagacaaa
gcgcggttggtatctaaacagagcattcgaaaagtgggaaaattaa

50

Sequence 1374

MHGLNGVEVTAKVKNVYRLVHSRRGAAKWVADVKAIDGKTWTIDNDFYSLPDENEENK
KTLYDKINHPSHYTYGEIEVIEFIEQVTKDYKPELAFAIGNAIIKYISRANRKNKEDLDK
ARWYLNRAFEKWEN*

55

Sequence 1375

Contig_0593_pos_3240_1561,

putative peptide of unknown function

atgaaaaaagtacaaaataaatattcaatcagaaagtttaacacagcagtaggtagtgtc

attgtagggacagcaattttctttggaggacaagcacacgctgcagaaaacgaagttcaa
 agagaacaacctaattgtagaacaagcagaacaacacagaagtacacacagacacacta
 caagcaagtaattgaagaagtgggtacaaaacaaccaagaaaaagaaacggccaaaatgat
 gtgtctacacaagcgacagagcaacctcgatttatcactaacacagatttcaaaacacaa
 5 acagacgaaaacggacaaacacctatttggagtaaaacaacagtcaattacgaatggaac
 gctacagggttataaaaaagggtgatgaaattaactttaacttctcctgaacagttaagactt
 gctaatagaacaaaactttgacttaaacacacctgataatgtcaatattggcagagttaat
 gcgacaagagatggattagtgatgtaagcttaactgatccgacagattacttagcgaca
 catgaaaataactaaagggttgatgtattttgagactatgttcaacagagataaagtcaaa
 10 gccggcgagagttacgacatcaaacttgggtgacaaggatcacagtagacgttgacaaa
 aatgaaatttaacaaaagtcggttacaaaaatggggatattgtagatgatgacaaaagta
 cgttgggatgtaagaataaaccaggatgaacaaactattaataatggacgtttagaagat
 acattgggtgacggtttaacattcgatgaagattcattaactgtcactgaattcgatgta
 gataatcaagagtttaggaagtcctttctatgattataaattaacaccaaccactaatgga
 15 tttaaatcgattttcttaaaacaaattaataaagcgatgaaattgaatacacacaacaa
 cctttattaggtacaaatcaccaatacacaaacagtggtgaattgactggtgacggatc
 aaagaacattagaaaatgttgaattctgaagtatctaacgctggtggaggtggtgaggtg
 gacaaacttccacctgtagagcctgaacaaacctacagaaccagagcaacctaaaggaact
 gaaactccggaagaaccaacaacaccaaacttccggaagaaccaaacacacctgaacaa
 20 ccgaacaatcctgaaacacctgaagaacctaacaaaccagaacaacctactaagtcagaa
 gaacaaaacaaccacgacctgaaacaccacaaactcctgaacaatcagaagttaaagaa
 aaacaccaagaacctaaaacaccaacagagaaaaaagaaacacctattacacctcaaaaa
 ccaagtaaaattgttgaggtagaaaataaagaagaagtgccacaaaagaatacaayat
 gacacgactttcgttgaacacgatcaaaagaacaacaaaacatattgataaaccagtt
 25 gaaagagttacgggtaacgtggctaatgaacaagaattagaaaaagagtcgaaagaagct
 gaaaaagtagaggaaaaagagcttccgaaaacaggacaagttgaaaatgtaggtgtcttt
 ggattgttagcactagtcactggtatcgacttgaagacgacgtaataaggaggattaa

30 Sequence 1376
 MKKVQNKYSIRKFNTAVGSVIVGTAIFFGGQAHAAENEVQREQPNVEQAETTQEVHTDTL
 QASNEEVVQNNQEKETAQNDVSTQATEQPRFITNTDFKTQTDENGQTPIWSKQVNYEWN
 ATGYKKGDEINFNLPEQLRLANEQNFDLNTPDNVNIGRVNATRDGLVNVSLTDPTDYLAT
 HENTKGWMYFETMFNRDKVKAGESYDIKLGDKGYTVDVAQNEINKSPLQKWGYVDDDNKV
 35 RWDVRINQDEQTINNGRLEDTLGDGLTFDEDSLTVTEFDVDNQELGSPFYDYKLTPTTNG
 FTIDFLKQINKAYEIEYTTTPLLGTNHQYTNSELVTDGDKYKLENVESEVSNAGGGGEG
 DNIPPVEPEQPTEPEQKPEPETPEEPTPNVPEEPNTPPEQPNNPETPEEPNKPEQPTKSE
 EPKQPRPETPQTEQSEVKEKHQEPKTPTEKKETPITPQKPSKIVEVENKEEVSPKEIQD
 DTFVVTTRSKEQPKHIDKPVERVTGNVANEQELEKESKEAEKVQEKELPKTGQVENVGVF
 40 GLLALVTGIALVRRRNKED*

Sequence 1377
 Contig_0594_pos_2850_3191,
 is similar to (with p-value 5.0e-39)
 45 >gp:gp|AF044951|AF044951_1 Staphylococcus aureus repressor p
 rotein (rzcA) and transport protein (rzcB) genes, complete c
 ds. NID: g3445565.
 atgagtgatattcatattaacgttgaaagcaatcaggatcacattgattaaagtcactcat
 attttcaaagctttaagtgattttaatcgtgtgagaataatggagtttcttgaaaacggt
 50 gaagcaagtggttgacatatttcacattctttaaatatgactcaatcaaatgtatcacat
 caattgaaactacttaaaagcactcatcttgttaaatctaaaagacaagggaatcaatg
 atttattctatagatgatatacacggtttcaactttacttaaacagccattcaccacgcc
 aaacatcctagtgaaagggtggaattttctaatgacaaatcataa

55 Sequence 1378
 MSGYHINVESNQDTLIKVTTHIFKALSDFNRVRIMEFLENGEASVGHISHSLNMTQSNVSH
 QLKLLKSTHLVKSQRQGSMIYSIDDIHVSTLLKQAIHHAKHPSEGGISNDKS*

Sequence 1379

Contig_0595_pos_1295_1999,
is similar to (with p-value 9.0e-42)
>gp:gp|AJ002481|LHAJ2481_1 Lactobacillus helveticus gene encoding transmembrane protein. NID: g3850046.

5 atggcattagtagtattttcgctcatatcaggtgcaggatgggcatttgggtcaaattattact
tttaaagcggttcgaattagtaggttcatcaagagcgatgccaaattactactgcatttcaa
ttacttgggtgcatctttatggggcggtttttgcgcttggaactggccccgtataacaaac
aaaatcattggatttctagctttactcgtaatccttataggtgcacgtatgactgtatgg
actgaaacaaagcaacaagaatatagtaaaaatctacgaagtgcagtgatcttattactt
10 gtagggtgaaattgggtattggatatattctgctgcacctcaagcaacggatattgggtgga
tttaaagcttttttacctcaagctataggaatgggtcattgtggctgtcatctatgcgttg
atgaatatgtctaaaggtaatgcttttaagagaaagtaagttggcaacaaacaatatcc
ggatttttctttgcgtttgctgctttaactattttaatttcagcacaacctaataatgaat
ggtttagcaacaggatttgttctatctcaaacatctgtagtattagcaacgctaacaggc
15 attttttcttaaatcagaaaaaacatcaaaagaattaatgattacaattgtgggatta
gttcttatttttagttgcagcatcaatcacagtgtttattaaataa

Sequence 1380

20 MALVFSLSISGAWFGQIITFKAFELVGSSRAMPITTAFQLLGASLWGVFALGNWPGITN
KIIGFLALLVILIGARMTVWTETKQQEYSKNLRSVILLVGEIGYWIYSAAPQATDIGG
FKAFLPQAIGMVIVAVIYALMNMSKGNFKEKVSQQOTISGFFFAAALTYLISAQPNMN
GLATGFVLSQTSVVLATLTGIFFLNQKTSKELMITIVGLVLILVAASITVFIK*

Sequence 1381

25 Contig_0595_pos_2021_2425,
is similar to (with p-value 2.0e-24)
>sp:sp|P44734|RBSD_HAEIN HIGH AFFINITY RIBOSE TRANSPORT PROT
EIN RBSD. >pir:pir|G64072|G64072 high affinity ribose transp
ort protein (rbsD) homolog - Haemophilus influenzae (strain
30 Rd KW20) >gp:gp|U32732|U32732_2 Haemophilus influenzae Rd se
ction 47 of 163 of the complete genome. NID: g1573480.
atgaagaaaacagcagtagtattaaatagtcacatttcaagcgcaatctccacactagggtcac
tatgatttattaacgattaatgatgcgggtatgcctatacctaataatgatgacaaacgtata
gatttagcagtgactaagtcattgccatgtttcattgatgtgttgagacagtggttaact
35 gaaatggaaatacaaaaaatatatttagcagaagaaattaaaactgcgaatgcacagcaa
ttaaaagcaattaaagaaattaatcaatgatgatgtagaaattaaatttattgcgcattct
gagatgaaagaaatgttaaaatctcctttaataaaggaaatatacgtactggtgaaatc
acccttttttctaactatctcctagaatctaattgtgacttttttaa

Sequence 1382

40 MKKTAVLNSHISIAISTLGHYDLLTINDAGMPIPNDDKRIDLAVTKSLPCFIDVLETVLT
EMEIQKIYLAEEIKTANAQQLKAIKKLINDDVEIKFIAHSEMKEMLKSPLNKGNI RTGEI
TPFSNIILESNTF*

Sequence 1383

45 Contig_0595_pos_3846_4859,
putative peptide of unknown function
atggaacgattttgtgtgtaaatcaaattaactatattcaaataaatccgtagaagcc
aaatttaaaacgagcgctctaagatcatggaaaactgatcaggcagatgctcataagcgtt
50 gcttgttttaggaccgagcgttaaacacagacaacttacctatacatgagttaatattc
tttgatttaagagaacgcgtccgttttcatctagaaatcgagaatgaacaaaatcgactt
aaatttcagatccttgaattactccatcaaacattccctgggttagaaagattatttagt
agtcgatattcaatcattgcactcaacatcgcaaaatctttactcatccagacatgggtt
cttgatatacgacaaggaggtactgattacacatatattcaattctacagataagggaatg
55 tcaatggataaagctacaaaatatgcacttcaattaagggtgatttgctcaagaaagctat
cctaattgctgatagacattcctttctagtcgaaaaattacgcttacttattcaacaattta
aaacaattctttcatcatctcaacaattagatgatgccatgattcaattagcacaacaa
ctcgattattttgaaaaatattcattcgatacctggatttggttaagctaagcacagctatg
attattggggagattggtgatattaagcgatttaaatcaataaacaactcaatgctttt

g ttggcattgatatcaaacgatatacaatcaggtcatcacactgtagagataccatcaac
 aagcgtggtaataaaaaagcgagaaaactttttattttgggtgattatgaatataataaga
 gggcagcatcattatgacaatcatgtcgtcgattattactacaaactaagaaagcagcct
 aatgagaaacctcataagactgccatcattgcttgataaatcgattattaaaaacaatt
 5 cattatcttgtaatgaatcataaattgtacgattatcaaatgtcaccacattag

Sequence 1384

MERFCCVNQINYIQMNPLEAKFKTSALRSWKTDQADAHKLACLGPTLKQTDNLPiHELIF
 FELRERVRFHLEIENEQNRLKFQILELLHQTFPGLERLFSSRYSI IALNIAEIFTHPDMV
 10 LDIDKEVLITHIFNSTDKGMSMDKATKYALQLRVIAQESYPNVDRHSFLVEKLRLLIQQL
 KQSIHHLKQLDDAMIQLAQQLDYFENIHSIPGIGKLSTAMIIGEIGDIKRFKSNKQLNAF
 VGIDIKRYQSGHTHCRDTINKRGNKARKLLFWVIMNIIRGQHHYDNHVVDYKYKLRKQP
 NEKPHKTAIIACINRLLKTIHYLVMNHKLYDYQMSPH*

15 Sequence 1385

Contig_0595_pos_5253_6743,

is similar to (with p-value 0.0e+00)

>sp:sp|P39211|XYLB_BACSU_XYLULOSE KINASE (EC 2.7.1.17) (XYLU
 LOKINASE). >gp:gp|U66480|BSU66480_19 Bacillus subtilis SpoVK
 20 (spoVK), YnbA (ynbA), YnbB (ynbB), GlnR (glnR), glutamine s
 ynthesase (glnA), YnaA (ynaA), YnaB (ynaB), YnaC (ynaC), Yna
 D (ynaD), YnaE (ynaE), YnaF (ynaF), YnaG (ynaG), YnaH (ynaH),
 YnaI (ynaI), YnaJ (ynaJ), xylan beta-1,4-xylosidase (xynB)
 , xylose repressor (xylR), xylose isomerase (xylA), xylulose
 25 kinase (xylB), YncB (yncB), YncC (yncC), YncD (yncD) and Yn
 cE (yncE) genes, complete cds. NID: g1750106. >gp:gp|Z99113|
 BSUB0010_55 Bacillus subtilis complete genome (section 10 of
 21): from 1781201 to 2014980. NID: g2634090.

atgggtgaaagaagtagttctaggaattgatttaggcacaagcgcaataaaaaattattgct
 30 gttgatcaactaggaaatgtcattgaatcagtaagcgaaacattaaagttataccaagag
 catcctggttatagcgaacaagaccctaataatggtttgaggctactaagaaagggata
 aaagaatttaattcaatcaacagaaatgtcagataagatagtaaaagggatttcttttca
 ggtcaaatgcattgggttggtcatagttgatgataatggcattcctttgagaaaagcgatt
 ttatggaatgatactagaaattcaatacaatgtagacaaattgaagatatatatggtgaa
 35 agattgaattacaatccgatattagaaggatttacacttcctaaaatggttatgggtacaa
 caacatgaacctgaaatttgaatcgagttgatgtttttatggtgcctaaagattattta
 cgttattgcttaacgcagacaattcatatggaatatagtgatgcattgtagtacattatta
 ttcaatcctgagaattatgaatggacaaaagatggttgagatacatttaacattggtgat
 atctatccacccttagtaaaatcacattcgtatggttggaatgtaacttcactcactggct
 40 aaagaattaggattatctagtgatggttgctgtatatgctgggggtggtgataatgcattg
 ggtgcaattggtgctggtgcatccatgataaaagtcattatgtagcataggtacttca
 ggtgttgattaaatggtgaataccaacgtgtgacctcatatgatagtaatttacactta
 ttcaatcatagtggtccagatacttattacgcaatgggagtaaacggttggcagcaggtcat
 agtttaaaactggttaaaaacaaacttttttgaaaatgaatcttttgaagagattttaaat
 45 ttagctgcatcttcaaagataggtgccaatggactactatttacacttacttagctgga
 gaacgtacgccacatggtgatgctcaaatacgtggaagttttataggtatcagtgaggcaa
 cataactaaagctgactttgcgagagcagtaaatcgaaggcataaacgtattctttatatgat
 tctataaaagattatgagacgagctggtcatgaaatgaactctatcacttcaatcggtggt
 ggtgctaagagtagattttggttacaacttcaagctgatatttttaatgtgcaataaaaa
 50 agattgaagcatgaagaaggcccaagcatgggagcggcaatttttagcggcatacgggtcta
 ggatggtttaaaacaattgagtccttggttagaggcatttattaaagtagacgaggtgtt
 gagccgaataatgaaaatcatgacctttatgaacaatactattcagtttatgaagctata
 tataaacaacgaacagcttactgctgatttggttaacgataaacgaattaa

55 Sequence 1386

MVKEVVLGIDLGTS AIKIIAVDQLGNVIESVSETLKLYQEHPGYSEQDPNEWFEATKKG I
 KELIQSTEMSDKIVKGISFSGQMHLVIVDDNGIPLRKAILWNDTRNSIQCRQIEDIYGE
 RLNYPNILEGFTLPKMLWVQQHEPEIWNRVDFMLPKDYLRYCLTQTIHMEYSDACSTLL
 FNPENYEWTKDVGDTFNIGDIYPPLVKSHSYGVNVTSSLAKELGLSSDVAVYAGGDNAC

GAIGAGVIHDKSALCSIGTSGVVLNVEYQRVTSYDSNLHLFNHVSVPDYYAMGVTLAAGY
 SLNWLKQTFEFENESFEEILNLAASSKIGANGLLFTPYLAGERTPHGDAQIRGSFIGISGQ
 HTKADFARAVIEGITYSLYDSIKIMRRAGHEMNSITSIGGGAKSRFWLQLQADIFNVQIK
 RLKHEEGPSMGAAILAAYGLGWFKTIESCVEAFIKVDEVFEPNNENHDLYEQYYSVYEAI
 5 YKQTKQLTADLLTITN*

Sequence 1387

Contig_0595_pos_2400_2083,

putative peptide of unknown function

10 atggttagaaaaaggggtgatttcaccagttacgtatatttcctttatttaaaggagatttt
 aacatttctttcatctcagaatgcgcaataaatttaatttctacatcatcattgattaat
 ttcttaattgctttttaaattgctgtgcattcgagttttaaatttctctgctaaatatatt
 ttttgattttccatttcagttaacactgtctccaacacatcaatgaaacatggcaatgac
 ttagtcactgctaaatctatacgtttgtcatcattaggtataggcataccgcgcatcatta
 15 atcggttaataaatcatag

Sequence 1388

MLEKGVISPVRIFFLPFKGDFNISFISECAINLISTSSLINFLIAFNCCAFVLISSAKYI
 FCISISVNTVSNNTSMKHGNDLVTAKSIRLSSLGIGIPASLIVNKS*

20

Sequence 1389

Contig_0596_pos_5465_6574,

putative peptide of unknown function

25 atggttaaacagtaaatgatattgtttctattgttatttagtgatattacacgtccaacgccc
 aaccatattcttgtagctttactaattgaggaattaaatcatgttcctcgtgagaatttc
 gtaattatttaattggtacagggactcatcgagatcaaacgcgagatgaattgattcaaatg
 ttaggtgaagatattgtaaatcagtaaaaaatcggtcaccaatcattgctcagaaaaagaa
 agtctagctaaagtgggacacagtcattatggatgtgatgtttatttaaacaagcatat
 gtagaatccgatttttaaaattgtaacaggttttattgaaccacacttttcgcccggattt
 30 tcaggtggacctaagggataatgcctggaattgcaggtttagaacaattcaaacattt
 cataatgcaaaaatgattggcgatccgagatcaacgtggggaatttagaagacaatcca
 gttcaagatatggcacgggaagttaaccgtatgtgtaaacctgactttttacttaattgtt
 gcattgaataaaagtaaaagaaattactgcagcatttgctgggtgaaatcttagatcacac
 aaagaaggatgagcatatgtaaaagatcatgcaatgtttaaatgtgagcaacgctttgat
 35 attgttatcgcatcaaattctggctatccttttagatcaaaatttatatcaaacagttaaa
 gggatgagtgagcagcagtaaaagttgttaaaaaagacgggtcatattattatggtatctgag
 tgtgcagatggctttcctgatcatggtaagtttgccgaaattttcaaaatggcagacaca
 cctcaaggtattttagaacttattcacaatccaaactttaaggaagttgaccaatggcaa
 gtacaaaaaacaagcaagtattcaaaacttttgccaatgtgcagtttattcagaacttact
 40 gaccaacaacttaagactcagatgttaatcccaacctctaacattgaacatacaatacaa
 gaattagaacatcgatatggccgtaaatcaaccattggtgttatgccacaaggctccttta
 acaataccgtacgtagaagataaagaataa

Sequence 1390

45 MVNSNDIVSIVISDITRTPNHILVPLLIEELNHVPRENFVIINGTGTHRDQTRDELIQM
 LGEDIVNSVKIVHNHCSEKESLAKVGHSQYGCDEVLYNKAYVESDFKIVTGFIEPHFFAGF
 SGGPKGIMPGIAGLETIQTFHNAKMIGDPRSTWGNLEDNPVQDMAREVNRMCKPDFLLNV
 ALNKSKEITAAFAGEILDTHKEGCAYVKOHAMFKCEQRFDIVIASNSGYPLDONLYQTVK
 GMSAASKVVKDKGHIIMVSECADGFPDHGKFAEIFKMADTPQGILELIHNPNFKEVDQWQ
 50 VQKQASIQTFFANVHVYSELTDQQLKDSMLIPTSNIEHTIQELEHRYGRKLITGVMPQGPL
 TIPYVEDKE*

Sequence 1391

Contig_0596_pos_6618_0,

55 is similar to (with p-value 1.0e-17)

>sp:sp|P73846|YH17_SYNY3 HYPOTHETICAL 30.2 KD PROTEIN SLR171
 7. >gp:gp|D90910|D90910_32 Synechocystis sp. PCC6803 complet
 e genome, 12/27, 1430419-1576592. NID: g1652956.

atgaaattagacgcactattgaaagacatgcagagtgtagtaattgccttctcaggtgga

gtagatagtagcttggtactgaaaaagcgattgatatttttaggtgtaactatgttaaa
 cctggtgtagtaaaatcagaattatttagaaatgaagagtttgactagcgcttaaaactt
 ggacaaagtctaggtggtgaagtattagaaactgaaatgtctgaacttcaagatgcgaat
 atcgtaaaaaatcgcctgaaagttggtactatagcaagcgcttgatgtatagtcaactt
 5 gagaatattaagaataaaactaggatttaattatgtgctagatggatgattatggatgac
 ttagatgattttcgtcccgattaaaagcaagagacgactttggtgttcgtagcggttta
 caagaagcaaaactatcttttagagcacagtggcgatgatatc

Sequence 1392

10 MKLDALLKDMQSVVIAFSGGVDSSLLKKAIIDILGVNYVKPVVVKSELFNNEEFELALKL
 GQSLGVEVLETEMSELQDANIVKNTPEWSYYSKRLMYSQLENIKNKLGFNYVLDGMIMDD
 LDDFRPGLKARDDFGVRSVLQEAKLSLEHSGDDI

Sequence 1393

15 Contig_0596_pos_3043_2615,
 putative peptide of unknown function
 gtgagtagtaaaattgaataaaaaatattaacattcaaaccgctcaagttttgaaacagaat
 ggtgaaaagcagagatttgagtttactacaaaagggttcttggaacaaaaatttgagat
 ttatacgttacgaagaacaaattgaagatgctaaagttaattgttacgattaaaaattgaa
 20 gatagcgggtgtaaagttaattcgtaaaggcgacattaatatgaacttacatttcgtcgaa
 ggacatgagacgacaacactctatgatgtacctaccggaaaaatacctttaactgttaaa
 aactaagccttatgcatttcgttactcataatggcggtaaacttaaaatacattatgag
 ttatatcaagatgaacaaaagatgggttcttatcaatatgaaataaattataaggagata
 agcgaatga

Sequence 1394

30 VSSKLNKNINIQTRQVLKQNGEKQRFEEFTKGSWQQKFADFIRYEEQIEDAKVNVTIKIE
 DSGVKLIRKGDINMNLHFVEGHETTTLYDVPTGKIPLTVKTLSLMHFVTHNGGKLKIHYE
 LYQDEQKMGSYQYEINYKEISE*

Sequence 1395

Contig_0596_pos_1862_957,
 is similar to (with p-value 0.0e+00)
 >gp:gp|Z99123|BSUB0020_30 Bacillus subtilis complete genome
 35 (section 20 of 21): from 3798401 to 4010550. NID: g2636240.
 >gp:gp|Z97024|BSZ97024_4 Bacillus subtilis ywiA, sbo, ywiB,
 argS and narK genes. NID: g2224752.
 gtgcacttttgataattggttagcgaaacatctttatatgaaaatggcgcgattaaaaat
 acattatctaaaatgaaagagctatagctatgaagcagatggcgcgacttggtta
 40 cgtacaagtgactttaaagacgataaagatcgtgtattaataaaaaaagacggtaattac
 acttatttcacaccagatacggcctatcactacaataagattaatagaggaaatgatatt
 ttaattgatattaatgggtgctgatcaccacggttatattaaccgtctcaaagctagctta
 gaaacatttggtgtagacagcgatcgtcttgaaattcaaattatgcaaattggttcgcctt
 atgcaaaacggagaagaagtgaagatgagtaagcgtacaggcaatgcgattactttacgt
 45 gaaattatggatgaagtgcgtatcgatgctgcacgttatttcttaacaatgcgtagtcct
 gattctcattttgactttgatttagagtttagctaaagagcaatctcaagacaatccaatt
 tattacgcacaatatgcacatgcacgaatttggttctattctaaagcaagctaaagaacaa
 ggtattgaagtgtctactgatgcagacttttctaaaataaataatgataaagcaatagac
 ttattaaaaaaagtagcagaattcgaatcgacgatagagagtgcagccgaacatcggtgca
 50 cctcatcgtctaaactaattatattcaggatctagctgcagcattccataaattttataat
 gctgaaaaagtgccttacagatgatacggaaaaaactaaagcacatgtagctatgattgaa
 gcggtgcgtattaccttgacaatgcattagcattagtaggtgttacagcaccagaatct
 atgtaa

Sequence 1396

55 VHFDNWFSETSLYENGAIKNTLSKMKELGYTYEADGATWLRTSDFKDDKDRVLIKDGNY
 TYFTPDATYHYNKINRGNDILIDLMGADHHGYINRLKASLETFGVDSRLEIQIMQMVR
 MQNGEEVKMSKRTGNAILTREIMDEVGIDAARYFLTMRSPOSHDFDLELAKEQSQDNPI
 YYAQYAHARICSILKQAKEQGI EVSTDAFDSKINNDKAIIDLKKVAEFESTIESAAEHRA

PHRLTNYIQDLAAAFHKFYNAEKVLTDDTEKTKAHVAMIEAVRITLHNALALVGVTAPE
M*

Sequence 1397

5 Contig_0596_pos_745_65,
putative peptide of unknown function
atgcgagatataacaatcatgcgaagattttagtaggggagaaatcatgttatcaatagaa
aaattatatcaaatctctatatcaaaatatgggccctcaatattggtgcccagcagaaaacg
ccaatagaaatgatgcttggggcaattctagtcctaaataactaattggaacaatgcagat
10 atagcgttatcaagattaaaagaagaacttcttttaatgcacagacgatattgaaaatg
cctttagaatcggttcagcaagtgatcggttcgagtggtttctataaaaaataaagctaag
gctatacaggcattgttactatggttaaatacaacatcattttgattatagtagtatagct
aagttatacgggtgatagcttaagaaaagaattactcaccatccgtggtataggtgaagag
accgccgatgtcttaatagtatatatttttaaggtaaagaattcatacctgatagttat
15 actagacgtatttttagaaaattgggatatacaatacagaaagttatcataaattgaaa
caggaattaacacttctgaatcattttcaaatcaagatgcaaatgagtttcacgcttta
ttagataattttgggaaaaattatttttaattggttaaggggaaacaacgctataacctttta
gatacctattttaaaaataa

Sequence 1398

20 MRDITIMRRFVGGIEIMLSIEKLYQILYQNMGPQYWPAETPIEMMLGAILVQNTNWNAD
IALSRLKEETSFNAQTILKMPLESQQVIRSSGFYKKNKAKAIQALLLWLNQHHFDYSSIA
KLYGDSLRLKELLTIRGIGEETADVLIVYIFKGKEFIPDSYTRIRIFRKLGYQHTESYHKLK
QELTLPEFSFSNQDANEFHALLDNFGKNYFNGKKGQRYTFLDTYFKK*

Sequence 1399

Contig_0598_pos_7659_9626,
is similar to (with p-value 6.0e-34)
>sp:sp|P37710|ALYS_ENTFA AUTOLYSIN (EC 3.5.1.28) (N-ACETYLMU
30 RAMOYL-L-ALANINE AMIDASE). >pir:pir|A38109|A38109 autolysin
- Enterococcus faecalis >gp:gp|M58002|STRHYDROLA_1 Streptoco
ccus faecalis bacterial cell wall hydrolase gene, complete c
ds. NID: g153658.
atgaagaaaaataaatttttagtatatttactatcgacggcgcttatcacgccaaaccttc
35 gctacacaaacagcttttgcgtgaagattcatctaataaaaaatacaaatcagataaaatg
gaacaacatcaatcacaaaaagaacatcaaaacaatctgaaaaagatgaatttaacaac
gatgattctaacaacagattctgatgataaaaaaagcacttctgacagcaaggacaaagac
tctaataaaccattatcagctgactcaacacatcgtaactataaatgaaagatgataat
ttagttgatcaactttatgataattttaagtctcagtcagtagatttttctaataactgg
40 gaaccgaataaatacgaagacagtttttagtttaacgtcactcatacaaaatttattlgat
tttgattctgatataacagattacgaacagccacaaaagacaagccattcttctaattgac
gaaaaagatcaagtagaccaagcagatcaggcaaaaacaacatcacacatcaagaacaa
tcacagtcgtctgctaacaagatcaagaatcatcaaacgatgaaaaagaaaaagacaact
aaccatcaagccgattctgacgtcagtgatttacttggagaaatggataaagaagatcaa
45 gaaggcgaaaacgttagatacaaaaacaaaatcaatcttcttctgagcaacaacaaactcaa
gcgaatgatgatagctcagaacgtaacaagaaatattctagtattacagattcagcatta
gactctatattagatgaatatagtcaggacgctaagaaaaacagaaaaagattacaataag
agcaagaatacaagtcacactaaaacatctcaaagtataatgccgacaagaatccacaa
ttaccaacagatgatgaattaaaacatcaatcaaaacctgcacaatcatttgaggatgac
50 attaaacgctcaaatacacggttcaacaagtcttttccaacaactacctgaattagacaat
ggtgacttattcttgattcatttaattgttggtagcagtcagacacacgtagtttcatt
caatcaattgctaaagatgcgcattcagattggaaaagaccaagatatatatgcatcagtt
atgattgctcaagctattttagaatctgactctggaaaaagttcacttgcacaatcacca
aatcataacttgtttggaatcaaaggtgactacaaaggacaatctgtaacttttaatact
55 ttagaagctgatagcagtaatacatatgttcagtatccaagcaggtttccgtaaatacca
agtactaaacaatctcttgaagattatgcagatttaatacaacatgggatcgatggtaat
ccgtcaattttataaaaccaacttggaaagtggaagctctatcatataaagatgctacttca
catctgtcacgctcatacgccacagatcctaattatttcaaaaaattaaatagtattatt
aaacattatcatttaacatcttttgcaaaagaaaaaatgcctaacatgaagaaatacaac

5 aatcaataggtacggatgtctctggtaatgacttcaaaccatttactgaaacttccggt
 acatcaccttaccacatggccaatgtacttggatgtgtaccaccgtatgaatcaattt
 gatgcatccatttctggtgacttaggtgatgctcataattggaataatcgtgctgaaagt
 gaaggctatacggtaacgcacacacctaataatcatactgcagttgtgtttgaagctgga
 10 caattaggtgctgatacacagtatggtcatgttgctttcgttgaaaaagttaatgacgac
 ggttcaattgttatttctgaatcaaatgttaaaggattaggtgtcatttcattcagaact
 attgatgcagaagatgctcaagatttagattacattaaaggtaaatag

Sequence 1400

10 MKKNKFLVYLLSTALITPTFATQTAFEDSSNKNTNSDKMEQHQSQKETSQKSEKDEFNN
 DDSKHDSDDKKSTSDSKDKDSNKPLSADSTHRNYKMKDDNLVDQLYDNFKSQSVDFSKYW
 EPNKYEDSFSLTSLIQNLFDSDITDYEQPQKTSHSSNDEKDQVDAQAKQPSQHQQEQ
 SQSSAKQDQESSNDEKEKTTNHQADSDVSDLLGEMDKEDQEGENVDTNKNQSSSEQQQTQ
 ANDSSERNKKYSSITDSALDSILDEYSQDAKKTEKDYNKSKNTSHTKTSQSDNADKNPQ
 15 LPTDDELKHQSKPAQSFEDDIKRSNTRSTSLFQQLPELDNGDLSSDSFNVVDSQDTRDFI
 QSIKDAHQIGKDQDIYASVMIAQAILESDSGKSSLAQSPNHNLFGIKGDYKQGQSVTFNT
 LEADSSNHMFISIAGFRKYPSTKQSLLEDYADLIKHGIDGNPSIYKPTWKSEALS YKDATS
 HLSRSYATDPNYSKKLNSIIKHYHLTSFDKEKMPNMKKYNKSIGTDVSGNDFKPFETETSG
 TSPYPHGGCTWYVYHRMNQFDASISGDLGDAHNWNRAESEGYTVTHTPKNHTAVVFEAG
 20 QLGADTQYGHVAFVEKVNDDGSIVISESNVKG LGVISFRTIDAEDAQDLDYIKGK*

Sequence 1401

Contig_0598_pos_7371_6229,
 is similar to (with p-value 3.0e-50)
 25 >sp:sp|P49022|PIP_LACLA PHAGE INFECTION PROTEIN. >gp:gp|L146
 79|LACPIP_1 Lactococcus lactis pip and gerC2 genes, complete
 cds's, and rrg gene, 5' end of cds. NID: g308860.
 atgaaaaacgcactaaaactttttatcacggatttaaaaagagttgctaaaacaccagggt
 gtatgggtcatcttagctggttttagcaattcttcttcattctatgcatggtttaacctc
 30 tgggctatgtgggatccgtatgggtcatcacaggacatatcaaaagttgccgtagtgaatgaa
 gaccaaggtgaaaaagttcgtggtaagaatattaatgtaggaaaaaaaatggtcaaaact
 ttaaaaaagaatgatagttttgactggcaatttgtgagtagagaaaaagccgaccatgaa
 attaagatgggaaaaatattatgcaggtatttatataccgaagaaattcacacatgaaatc
 actggtacttttaagaaaaacatcctcaaaaggcggatatagattttaaagtaaatcagaag
 35 attaatgctgtagcagctaagttaaccgatacgggatcgctcgtttgtgattgataaagca
 aataaacaatttaacaaaaccgtagcaaccgctttactttctgaagctaataaagtcgga
 ctatcaattgaagataatgtacctacaatcaataaaaattaagagtgctgtatatcaagct
 aataattcattgcctaaaattaatcaatttgcagacaagattattgaactaaaataaacat
 caagcagatttggatgcttatgctaataatcatttagaagtttaggaaagtataaagggaat
 40 gtattagcgtcgaagaaaaacttaatgctgttaattcgtctattccggcgcttaatgaa
 agggctaaattgatacttgcaacttgatagctacatgcctaataattgaaagaattttaaat
 gttgctgctaataatgatgttccagcacaatttcctagaattaataggggtgtcgatattgca
 agtgaaggattatgatgcagcgagtggtcagttaaatgatgcaaaagggttatttgactcaa
 gctaaagcgagagtgaggagactatcaagaagcagctggccgcgctcaagatgtgaacaac
 45 caagcaaatcaaaatctaagaaatcaaacatcaactacaccccaaagcgctataaaatca
 tcgcattcgggaagggaagagtcattcaagcattaaaacagtacctgtgagtcattcttgc
 taa

Sequence 1402

50 MKNALKLFIIDLKRVAKTPGVWVILAGLAILPSFYAWFNLWAMWDPYGHIGHIKVAVVNE
 DQGEKVRGKNINVGKMKVKTLLKKNDSFDWQFVSREKADHEIKMGKYIYIIPKKEFHEI
 TGTLRKHPQKADIDFKVNQKINAVAAKLTDTGSSFVIDKANKQFNKTVATALLSEANKVG
 LSIEDNVPTINKIKSAVYQANNSLPKINQFADKIIELNKHQDDLDAYANQFRSLGKYKGN
 VLDAQEKLNAVNSSIPALNERAKLILALDSYMPNIERILNVAANDVPAQFPRIINRGVDIA
 55 SEGIDAASGQLNDAKGYLTQAKARVGDYQEAAGRAQDVNNQANQNLRNQTSTTPQSAIKS
 SHSEGKSHSSIKTVPVSQSC*

Sequence 1403

Contig_0598_pos_2393_1767,

is similar to (with p-value 4.0e-27)
>gp:gp|U51115|BSU51115_13 *Bacillus subtilis* CotA (cotA), GabP (gabP), YeaB (yeaB), YeaC (yeaC), YebA (yebA), GMP synthetase (guaA) genes, complete cds, and AIR carboxylase I (purE) gene, partial cds. NID: g2239287. >gp:gp|Z99107|BSUB0004_88 *Bacillus subtilis* complete genome (section 4 of 21): from 600701 to 813890. NID: g2632866.
atgtcagcaattgctcaaaatccatgggttaattggtcttagcaattttataattaatgtatgttatgtcacatttttaacaatgagaacgatattgactttgaaagggtatcggttatgttgcagcagtagttagttttatggaagtcttagtttatgttggttagggttagtaattcttagcctagaccaaataccaaatattttgcttatgcattaggattctcagtcggtattatagtaggaatgaaaatcgaggaaaaacttgcggttaggttatcacagttgtcaatgttacttcacggaatattgagtttagatttaccaaatgaattaagaaatttaggggtatggtgtaaccactacgaagcattttggtagagatggaagtccaatggaatgcaaatattaacaccaaga
aaatatgaattaaagttaatggatactgtcaagaacttagatctcaaggcattttattatagcgtatgaaccaagaaatattcacggaggattctgggttaagggtgtacgtaaacgtaaa
ctgaaagcttatgaaccagaacaactggaagttgtagtagatcacgaagaaatagtaggtgtagctcaaataagcaaaaagttagt

Sequence 1404
MSAIAQNPWLMVLAIFIINVCVVTFILTMRTILTLKGYRYVAAVVSFMEVLVYVVGGLGLVMS
SSLDQIQNIFAYALGFSVGIIVGMKIEEKLALGYTVVNVTSSEYELDLPNELRLNLGYGVTH
YEAFGRDGRMVMQIILPRKYELKLMDTVKNLCLKAFIIAYEPRNIHGGFWVKGVRRKRL
LKAYEPEQLEVVVDHEEIVGGSSNEQKV*

Sequence 1405
Contig_0598_pos_0_1243,
is similar to (with p-value 0.0e+00)
>sp:sp|P12047|PUR8_BACSU ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL). >pir:pir|C29326|WZBSDL adenylsuccinate lyase (EC 4.3.2.2) - *Bacillus subtilis* >gp:gp|J02732|BACPURF_3 *B. subtilis* pur operon encoding purine biosynthesis enzymes, 12 genes. NID: g143363. >gp:gp|Z99107|BSUB0004_92 *Bacillus subtilis* complete genome (section 4 of 21): from 600701 to 813890. NID: g2632866.
atgtacagttttttaggagctataatacatacgataggagtttacaagaatgatagaa
cgattattcaagagatgaaatgtctagattttggacggatcaaaatcgctatgaagcatggttagaagtagagattctagcgtgtgaagcatggagtgagtttaggttatattcctaagaa
gatgtaaaaaaaaattcgtgaaaatgccaaaggtaaatggtgaacgcgcgctaagaaattgaa
caagaacacgcacacgatgtagttgcatttacacgcgcaagtttctgaaacttttaggtgat
gaacgtaagtgggtgataacttatggaattccactcatgttagttgacactgccttaagt
tatgttaattaaacagcgaacgaatccttgaaaaagatttagaaagatttatagatgta
ttagctgccaagctaaaaaatatcaatatacattgatgatggggcgtaacacatggagtt
catgctgagcctacaacttttggcgtaaaaatggctcctttggtatactgaaatgaaacgt
aatcctaagcggttttaagaagtgcgtaagaaatagaagtagggaaaatgagtggaagca
gtaggaacatttgccaatattccaccgaaattgaggcttacgtatgtgaacatttagga
atagacactgcagctgtttctactcaaacattacaaagagatcgctcatgcttattacatt
gctacatagcactaattgcatcatcaactggaaaaatttgagtagaaatcgtaattta
caaaaaactgaaactagagaagtcgaagagcatttgcaaaaggacaaaaaggatcatca
gcaatgccacataaaagaaatccaatcggttctgaaaatattacaggtatatctcgtgtg
attcgaggatatattactactgcatacagaatatacctttatggcatgaaagagatatc
tctcattcttctgcagaacgtatcatgttaccagatgttacaattgctctagattatgca
cttaatcgatttacgaatattggtgatcgactaacagtttatgaagataatatgagaat
aatattgataaaacgtatggttttaattctctcacaacgcgctactactagcattaattaat
aaaggcatggttagcgaagaagcttatgataaagttcaacctaaagcaatggaattcttg
gaaacaaaaaacaccatttagaagacttaattgaacaaagattcat

Sequence 1406
MYSFCRSYNTYDRSLQRMIERYSRDEMSSIWTDQNRYEAWLEVEILACEAWSELGYIPKE

DVKKIRENAKVNERAKEIEQETRDVVAFTTRQVSETLGDERKWVHYGLTSTDVVDITALS
 YVIKQANEILEKDLERFIDVLAACKAKKYQYTLMMGRTHGVHAEPTTFGVKMALWYTEMKR
 NLKRFKEVRKEIEVGMMSGAVGTFANIPPEIEAYVCEHLGIDTAAVSTQTLQDRRHAYYI
 ATLALIATSMEKFAVEIRNLQKTETREVEEAFKQKQSSAMPHKRNPIGSENITGISRV
 5 IRGYITTAYENIPLWHERDISHSSAERIMLPDVTIALDYALNRFNIVDRLTVYEDNMRN
 NIDKTYGLIFSQRVLLALINKGMVREEAYDKVQPKAMESWETKTPFRELIEQDSX

Sequence 1407

Contig_0599_pos_1311_1619,
 10 putative peptide of unknown function
 atgatggccaacacctttaataataaccaatacatTTTTccatacagagcggttcggttcatt
 ccgcgtgataatagtaatgcagttaaaataatcactacagcagcaatgatataatgaca
 ccaccgttacttccaaatggattagataatgatttaggttaaagaatgccaatggtgca
 ataagacctcttaagtttagcagaaaagcctgaagcaacgaaagcaacagcaataaagtat
 15 tctgctaaaagcgcccaaccggcaacccatccgaataattcaccaaaaaagaacattaatc
 catgaataa

Sequence 1408

MMANTFNITNTFSIRAASFIPRDNNAVKIITTAAMISMTPLLPNGLDNDLGKEMPNGA
 20 IRPLKLAEKPEATKATAIKYSAKSAQPATHPNNSPKRTLIHE*

Sequence 1409

Contig_0599_pos_2684_3322,
 is similar to (with p-value 8.0e-40)
 25 >sp:sp|P39592|YWBI_BACSU HYPOTHETICAL TRANSCRIPTIONAL REGULA
 TOR IN EPR-GALK INTERGENIC REGION. >pir:pir|S39679|S39679 hy
 pothetical protein - Bacillus subtilis >gp:gp|X73124|BSGENR_
 25 B.subtilis genomic region (325 to 333). NID: g413923. >gp
 :gp|Z99123|BSUB0020_126 Bacillus subtilis complete genome (s
 30 ection 20 of 21): from 3798401 to 4010550. NID: g2636240.
 atgaatgacattgtgaacgttcaaaaaggtcatattaaaataggcttatcaccaatgatg
 aatgttcaaatgtttacaaatgcattgaatcagtttcacagactctatcctaattgtgaca
 tatgaagtgattgaggggtggtggttaaaattgttgagaacttaacatctaattgatgatgtg
 gatattggtattactacattacctgtagatcacactgaatttcattcaacttctttatat
 35 aatgaagaattattattagtagtaagtaatgaccatcatttagcacatttaataaagta
 gacatggcagatttgaaagatgaagagtttgtttatttcatgatgattattattaaaa
 gatcaaatatagagaactgtaaaaggctaggctattaccctaaaactgttgctaatt
 tctcaaatagttttatcgctaataatgattcaacaaggaataggaattagtagtgcgtcca
 gaaagtttagttaatttaattggggaataacgtaacgtccattcaattagagaatgttgaa
 40 ttatcatggcatcttggcgtgatatggagaaaagatgcttatctcaatcatgtaactcgc
 aatggattgaattttatttctgagatgaaaccaacatag

Sequence 1410

MNDIVNVQKGHIGLSPMMNVQMFTNALNQFHRLYPNVTYEVIEGGGKIVENLTSNDDV
 45 DIGITLTPVDHTEFHSTSLYNEELLVVSNDHHLAHLNKVDMADLKDEEFVLFHDDYYLK
 DQIIENCKRLGYYPKTVANISQISFIANMIQQGIGISIVPESLVNLMGNVNTSIQLENVE
 LSWHLGVIWRKDAYLNHVTRKWIEFISEMKPT*

Sequence 1411

Contig_0599_pos_4138_5106,
 50 is similar to (with p-value 8.0e-89)
 >pir:pir|A25805|A25805 L-lactate dehydrogenase (EC 1.1.1.27)
 - Bacillus subtilis
 atgaaggagttcggttaaaatgaaaaaatttgggaaaaaagttgttttagtaggagacggt
 55 tccgtagggttcaagttatgcatttgcattggtgactcaaggaattgcagatgaatttga
 attattgatattgcaaaagataaagtgggaagcagacgttaaagatttaaacatggtgca
 ctttacagttcttccaccagtgaactgtaaaagctggagaatatgaagattgtaaagatgca
 gatttagttgttattacagcaggtgcacctcaaaaaccgggtgaaactcggttacaactt
 gttgagaaaaataactaaaatcatgaaaagtatcgtaactagtgatgcatgtagtggttt

gatgggttcttccctaattgctgcaaaccagttgatattcttaacacggttatgttaazagaa
 gttacaggtttaccagctgaacgtgttattggttctggtacagtgccttgatagtgcaga
 ttcatagatttaataagtaaagaattaggtgttacatcaagtagtgttcacgctagcatt
 ataggtgaacatggtgactctgaacttgcaagtttgggtctcaagcaaacggttgagggtatt
 5 tcaagtgtatgatacattgaaagaagaactggtagcgatgctaaagcgaatgaaatttat
 attaatacaagagatgctgcttacgatattcattcaagctaaaggatctacgtattatggt
 atagctctagcactattacgtatttctaaagctttactaaataatgaaaatagtattttg
 acagtttctagtcaacttaattggtcaatatggatttaacgatgtttatcttggttacca
 acacttatcaatcaaaatggtgcagttaaaaatttatgaaacaccattaaatgataacgaa
 10 ctacaattactagaaaaatcagtgaaaactttagaagacacttatgattctataaaacat
 ttagtttaa

Sequence 1412

MKEFVKMKKFGKKVVLVGDSVGSYAFAMVTQGIADFEVVIDIAKDKVEADVLDNLHGA
 15 LYSSSPVTVKAGEYEDCKDADLVVITAGAPQKPGETRLQLVEKNKIMKSIIVTSVMDSGF
 DGFFLIAANPVDILTRYVKEVTGLPAERVIGSGTVLDSARFRYLISKELGVTSSSVHASI
 IGEHGDSELAWSQANVGGISVYDTLKEETGSDAKANEIYINTRDAAYDIIQAKGSTYYG
 IALALLRISKALLNNENSILTVSSQLNGQYGFNDVYLGLEPTLINQNGAVKIYETPLNDNE
 LQLEKSVKLTEDTYDSIKHLV*

20

Sequence 1413

Contig_0599_pos_5169_6833,

is similar to (with p-value 0.0e+00)

>gp:gp|L16975|LACALS_2 Lactococcus lactis alpha-acetolactate
 25 synthase (als) gene, complete cds. NID: g473900. >gp:gp|A23
 961|A23961_1 L. lactis alpha-acetolactate synthase gene. NID
 : g809617.

atggcggaaaaacaatatctgcagcacaaatggtaattgatactttaaaaaataatgga
 gttgagtagtatttgggtattccaggtgcgaaaatcgactacttatttaatgcactagag
 30 gatgacgatattgaattagtcgttacgcgtcatgaacaaaacgcagcgatgattgcacaa
 ggtattggtcggtttaacagggaaaaccaggtgtggctattactacaagtggccaggggta
 agtaacttaactactggtttatttaactgcaactctctgaaggtgacctgtattagctatc
 ggtggtcaagttaaaagaaatgacttattacgtttaacacatcaaagtattgataacgca
 tcattacttagatcctcaactaaatatagtcgagaagtacaagatccagaatcactatca
 35 gaggttattacgaatgcaatgcgtacagccacttcaggtaaaaatggagcgagctttatc
 agtattccacaagatgttatttcatcacctgtcaaagctgatgcaatttcattatgtcaa
 aaaccacatcttgggtgtaccttcagaacaagaaatgaatgaagtgttgaggctattaaa
 aattctaaattcccagttatttagctgggaatgagaagttcaagccaagctgaaacagaa
 gctattcgccgttttagttcaaaaaacaaatttacctgttgtgaaacattccaaggtgcc
 40 ggcgttaattagtcgcaattagaaaaatcacttcttcgggtcgtgttggtttatttagaaac
 caagtgggtgacgaattacttagaaaaagtgatttagttatcacaatcggttatgacctt
 attgaatatgaagcaagtaactggaataaagaattagatactaaaatcatcaatgtcgat
 gaagaacatgctgaaatcactaattacatgcaaccagttaaagagttaatcggaacatt
 gcaggtacaatagatatgatttctgaacatgtaaatgaaccatttattaatcaagatcat
 45 ttagatgaacttgaaaaatgaagaggcgaaatcacagaagcaactggaattaaagcaact
 cacaaagaaggtgtgatgcaccagttgaaatcattgaaacaatgcaaaaagttttaact
 gatgatactactgtaactgtagatgtgggaagccattacatttggatggctcgtaaatc
 agaagttacaatcctagacatttactatttagtaacggtatgcaactctaggtgttgca
 cttccatgggctatttgcagctgcacttgtacgtccaaatcacaaagttgttctgtagct
 50 ggagacggtggtttcctattctcaggacaagaattagaaaactgcagtagctaaaaactta
 aatattcattcaattaatttggaatgatggtcgttataacatggttgaaattccaagaagaa
 atgaatatataacgctcttcaggtgtagaatttgaccagttgatttatgtaaaatgca
 gaatcatttggcgttaaaggattacgtgtgactaatcaagaagaattagaggcagcactt
 aaagagggttacgaaactgatggaccagttatttgaatcccagtttaactatgcagat
 55 aatgttaaattatctacaaatatgtttacaaatgcttttaattaa

Sequence 1414

MAEKQYSAAQMVIDTLKNNGVEYVFGIPGAKIDYLFNALEDDDIELVVTRHEQNAAMIAQ
 GIGRLTGKPGVAITTSGPVSNLTTGLLTATSEGDPLAIGGQVKRNDLLRLTHQSIDNA

5 SLLRSSTKYSAEVQDPESLSEVITNAMRTATSGKNGASFISIPQDVISSPVKADAISLCQ
KPHLGVPSEQEINEVIEAIKNSKFPVLLAGMRSSSQAEATEAIRRLVQKTNLPVVETFOGA
GVISRELENHFFGRVGLFRNQVGDELLRKSDLVITIGYDPIEYEASNWNKELDTKIINVD
EEHAEITNYMQPVKELIGNIAGTIDMISEHVNEPFINQDHLDELEKLARGEITEATGIKAT
HKEGVMHPVEIETMQKVLTDTTVTVDVGSYIWMARKYRSYNPRHLLFSNGMQTLGVA
LPWAISAAALVRPNTQVVSAGDGGFLFSGQELETAVRKNLNIIQLIWNDRYNMVEFQEE
MKYKRSSGVFEGPVDYVKYAESFGAKGLRVTNQEELEAALKEGYETDGPVLIDIPVNYAD
NVKLSTNMLFNALN*

10 Sequence 1415
Contig_0599_pos_6987_7649,
is similar to (with p-value 9.0e-54)
>gp:gp|U92974|LLU92974_24 Lactococcus lactis unknown gene, p
artial cds, and HisC (hisC), unknown, HisG (hisG), unknown,
15 HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF),
HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB),
LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB),
IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR)
genes, complete cds. NID: g2565137. >gp:gp|S82499|S82499_1
20 aldB=alpha-acetolactate decarboxylase [Lactococcus lactis, s
sp. lactis, NCDO2118, Genomic, 840 nt]. NID: gl699351.
atggcaggtttattagaaggaactgcttcaattaatgacttattagaacatggtgattta
ggtattgctacgcttacaggttctgatggagaagtaatttttggtagtggttaaagcttat
catgcaaatgaacataaagaatttatagaattgacaggcgacgaaatgacaccatagca
25 actgttcaaaaattcaaaagcagactcaagttttaaacaatctaataaaaatcaagaagaa
gtattcgatgaagttaaaaaacaatgaaaagtgaataatgttctcggcagttaaaatt
tcaggaacgtttaaaaaaatgcatgtacgtatgatgcctggtcaagaacctccatacaca
cgtttaattgattcagctcgtagacaacctgaagaacacgtgaaaatatcaaaagtttca
atcgtaggtttcttactccagaattattccatggtattggttctgcaggtttccatatt
30 cactttgcaaatgatgatcggtgattttggtggtcatattttagactttgaagtggatgat
gtgactgttgaaatacaaaaactttgaaacatttgaacaacacttcccagtagatgctaaa
tcatttactgatgctgacattgactataaagatatagccgatgaaatcagagaagctgaa
taa

35 Sequence 1416
MAGLLEGTASINDLLEHGDLGIATLTGSDGEVIFVDGKAYHANEHKEFIELTGDEMPYA
TVTKFKADSSFKTSNKNQEEVFDEVKKQMKSENMFSAVKISGTFKKMHVRMMPGQEPPT
RLIDSARRQPEETRENIKGSIVGFFTPELFHGIGSAGFHIHFANDDRDFGGHILDFEVD
VTVEIQNFETFEQHFPVDAKSFTDADIDYKDIADIREAE*

40 Sequence 1417
Contig_0599_pos_9020_9811,
is similar to (with p-value 1.0e-70)
>sp:sp|P52996|PANB_BACSU 3-METHYL-2-OXOBUTANOATE HYDROXYMETH
YLTRANSFERASE (EC 2.1.2.11) (KETOPANTOATE HYDROXYMETHYLTRANS
45 FERASE). >gp:gp|L47709|BACYP1A_17 Bacillus subtilis (clone Y
AC15-6B) ypiABF genes, qcrABC genes, ypjABCDEFGHI genes, bir
A gene, panBCD genes, dinG gene, ypmB gene, aspB gene, asnS
gene, dnaD gene, nth gene and ypoC gene, complete cds's. NID
50 : gl146223. >gp:gp|Z99115|BSUB0012_183 Bacillus subtilis com
plete genome (section 12 of 21): from 2195541 to 2409220. NI
D: g2634478.
atgaaggcatcacagcaaaaagatttctatggttacagcttatgattatcctagtgctaag
caagcacaacaagctgaaattgacatgattttggttaggagattcttttaggaatgacagt
55 ttaggatatgatagtactgttcaagttacattgaacgatatgattcatcatggttaaggct
gttaaaagagggtgcttcagatacatattatagttggtgatagccatagggactgttggt
ttaagtgatgaagaagatctaaaaaatgcacttaagctttatcaaaacacgaatgctaac
gctgtcaagtagaagggtcctcatcttacatcatttattcaaaaagcaactaaaatgggt
atacctgttgtttctcacttaggtcttacacctcaaagtgtaggtgtaatgggtataaa

cttcaaggggatacaaagacagccgctatgcaacttatcaaagatgctaaagctatggaa
 actgctggtgcagtagtactggttttagaagccatacctagtgttagctcgagaaatt
 agtcagcaactcactattccagttataggtataggggcaggaaaagatactgatgggcaa
 gtgttagtgtatcatgatatgttaaattatggtgtgatcgacacgctaagtttgttaag
 5 caatttcagacttttcaagtggattgatggattaaggcaatataatgaagaagttaa
 gcaggagcgtttccttctgaaaatcatacttacaaaaaacgtattatggatgaggtagag
 caacatgactaa

Sequence 1418

10 MKASQQKISMVTAIDYPSAKQAQQAIDMILVGDSLGMTVLGYDSTVQVTLNDMIHHGKA
 VKRGASDTFIVVDMPIGTVGLSDEEDLKNALKLYQNTNANAVKVEGAHLTSFIQKATKMG
 IPVVSHLGLTPQSVGMVGYKLQGDTKTAAQMLIKDAKAMETAGAVVLVLEAIPSDLAREI
 SQQLTIPVIGIGAGKDTGGQVLVYHMLNYGVDRHAKFVKQFADFSSGIDGLRQYNEEVK
 AGTFPSENHTYKKRIMDEVEQHD*

Sequence 1419

Contig_0599_pos_8950_8048,
 is similar to (with p-value 1.0e-17)
 >sp:sp|Q50648|Y0BS_MYCTU HYPOTHETICAL 26.2 KD PROTEIN CY227.
 20 28C. >gp:gp|Z77724|MTCY227_6 Mycobacterium tuberculosis H37R
 v complete genome; segment 114/162. NID: g3261620.
 atgaatcgtaaaaaagaaaggaaagaataaagttatgacaattgccattataggccagggt
 gcagtggttacaacgttagcttttgaattaaaaaaagttctaccagatacggaaactcatc
 ggccggcaagataaaattaatgacctatttccagaaaaacttctaattggaagtaattgtt
 25 aaagtgcatttcatttaatacatattaaatcaaacttttgatgtcattatcatagcagttaaa
 acacatcaattggatgacgtcattaaacaattacctaataatcactcatgacgattcgctc
 attatcttagcacaaaaatggctatggacagcttaataaaacttccatatcaacatgtcttt
 caagcagtcgtctatattagcggacaaaaagttaacaacaatgttcaacatttcagagat
 taccaactatataattcaagatagcacactaaactcgtcaattcaagcaaatggttcacct
 30 tccaaaatagaggtggttttacaagaaaaattgaaaaaagcatttggtataaattatta
 gtgaatttaggtataaaataccatcactgctattggacaacaaccagctaaaaattttaaaa
 tctctcatatttgatgctgtgtgctgctgctatattagttgatggctttaaagtgtgctaga
 gctgaacaaattgactttgaagatcatatcggtgatgatattttaaatatttataaagggt
 tatccagacgaaatgggaacaagtatgtattacgatgtcattaacaagcatcctcttgaa
 35 gtcgaggccatacaggggttatatatataaatgtgcaaaaaaacatcatttagagacacc
 tatctagatatggcttatacatttttatacgcttatcaccttgaatacacacaaccagat
 tga

Sequence 1420

40 MNRKKKGKKNKVMITAIIGPGAVGTTLAFELKKVLPDTELIGRQDKLMTYFPENTSNGSNV
 KVTSENHINQTFDVIILAVKTHQLDDVIKQLPKITHDDSLIILAQNGYGQLNKLPHYQHV
 QAVVYISGQKVNNNVQHFRDYQLYIQDSTLTRQFKQMVHPSKIEVVLQENIEKSIWYKLL
 VNLGINTITAIQQPAKILKSPHIESLCRRILVDGLKVARAEQIDFEDHIVDDILNIYKG
 YPDEMGTSMYYDVINKHPLEVEAIQGYIYKCAKKHHLETPYLD MAYTFLYAYHLEYTQPD
 45 *

Sequence 1421

Contig_0599_pos_1915_743,
 is similar to (with p-value 4.0e-49)
 >gp:gp|AL023702|SC1C3_2 Streptomyces coelicolor cosmid 1C3.
 NID: g3169026.
 atggttatggaggtttgttatgggaagtttttttaatcggatgactcgaaaagagaatcct
 actatttatcaaagtaagatgggcatcttaacgcacattacgtgtacgcgactttctt
 gcactaggtgttggtacaattgtttctacatctatcttcaactttaccaggtgttgctcg
 55 gctgagcatgccggacctgctgtttcattgtcattcttattagctgctattgtggcaggt
 ctgtgtagcctttacttatgcagaaatggcatctacaatgccttttgctgggtcagcttat
 tcatggattaatgttcttttttggtgaattattcggatgggttgccggttgggcgctttta
 gcagaatactttattgctgttgctttcgttgcttcaggcttttctgctaacttaagaggt
 ctatttgcaccattgggcatttctttacctaataatcattatctaatacatttgaagtaac

gggtggtgctattgatatcattgctgctgtagtgattatlttaactgcattactattatca
cgcggaatgaacgaagccgctcgatggaaaatgtattggttatattaaggtgttgcc
atcattttatgtgattgttgggtaactgcgattaatttcagtaactatatacctttt
attccagaacataaagtactgaaactggcgactttggaggttggaaggtatttatgct
5 ggagtttcaatgatttttttagcttatattggttttgactctattgctgctaattcaqct
gaagcgattaatccacagaagacaatgcctagaggaatcttagggtcactcatagtagca
attgtattgtttgtggccgtagcacttgttcttgttggcatgttccactactctcaatac
gctgataatgcagagccagtaggttgggcattacgagaaaagtggtcatggtattattgct
gcaattgttcaagcaatttctgtcatcggtatgttccactgcattaatcggtatgatgctt
10 gcaggttcacgtctattatattcatttggacgagatggtttactcccttcttggttaagt
caattgaatcacaacatttacctaactcgagcacttgcatacttacaatcattggcgta
gttatcggtcaatgttccctagcaatcgctaa

Sequence 1422

15 MLWRFVMSFFNMRTRKENPTIYQSKDGHKLKRTLVRDRLALGVGTIVSTSIFTLPGVVA
AEHAGPAVSLSFLLAAIVAGLVAFTYAEMASTMPFAGSAYSWINVLFGELFGWVAGWALL
AEYFIAVAFVASFANLRGLIAPLGISLPKSLSNPFGSNGGVIDIIAAVVIILTALLS
RGMNEAARMENVLVILKVLAILFVIVGLTAINFSNYIPFIEHKVTETGDFGWQGIYA
GVSMIFLAYIGFDSIAANSAAEAINPQKTMPRGILGSLIVAIVLFVAVALVLVGMFHYSQY
20 ADNAEPVGWALRESHGIIAAIVQAISVIGMFTALIGMMLAGSRLLYSFGRDGLLPWSLS
QLNHKHLNRLVILTIIGVVIGSMFPSNR*

Sequence 1423

Contig_0603_pos_1490_1990,
25 is similar to (with p-value 7.0e-38)
>gp:gp|AF082668|AF082668_1 Streptococcus pyogenes CsrR (csrR
) and CsrS (csrS) genes, complete cds. NID: g3599370.
atgcttccaaacataaatggtctagaaattttagacaaaattcgtaaaaaaactact
ccaattatcatcattactgcaaaaagcgagacatatgataaagtagctgggttgactat
30 ggggcagatgactacattgtaaaaccctttgatatagaagaattgctcgcaagaataaga
gcggtattgcgagacagccagataaagatgttttagatatcaatggtattatcattgat
aaagatgccttttaagtactgttaattggccatcaattagaattaactaaaacagaatac
gatttattatatgttttagctgaaaatcgtaaccacgtcatgcagcgtgaacaaattctc
gatcacgtatgggggtataatagtgaagtagaaacgaatgtcgttgatgtttacattcgt
35 tatttacgtaataaaactcaaacccttttaataaagaaaaatccatagaacagtagctggc
gtagggtatgtgattcgatga

Sequence 1424

40 MLPNINGLEICRQIRQKTTTPIIIIITAKSETYDKVAGLDYGADDYIVKPFIDIEELLARIR
AVLRRQPKDKVDLDINGIIIDKDAFKVTVNGHQLELTKTEYDILLYVLAENRNHVMQREQIL
DHVWGYNSEVETNVVDVYIRYLRNKLKPFNKEKSIETVRGVGYVIR*

Sequence 1425

Contig_0603_pos_2503_3357,
45 is similar to (with p-value 2.0e-42)
>gp:gp|U81166|LU81166_1 Lactococcus lactis subsp. cremoris
MG1363 histidine kinase (llkA) gene, complete cds. NID: g2
182834.
gtgagttatatcttttcttcgcaaattactaaaccgatagttacaatgtccaataaaatg
50 aatcaaatagaagagatgggttttcaaaaataaacttgaattaactacaaattatgaagaa
acagataatttaattgatacttttaattgaaatgatgtatcaaatagaagaatcttttaatt
cagcaacgtcaatttgcgaggatgcttcacacgaattaagaacgccactgcagattatt
caaggtcatctaaatttaacccaacgttgggggaaaaaagatccagcagtttgggaagaa
tctttgaatatattcaattgaagaagtgaatcgaataacaaaacttgcgaagaactactt
55 ttacttaccaaagatagagtcaatcataatgttttggatgtgaaaatgtagacgtaaat
agcgagattcaatcacgtgtgaagtcactgcaacacctacatccagattatacttttgaa
acacatcttgctactaagcctatccaattaaaaattaaccgtcatcagtttgaacaactc
ttactcatattttattgataatgcaatgaaatcagacactgaacataagcattaaaatt
gttactcaactaaaaaataaaatgattatgattgatattactgatcatggtatgggtata

ccaaaagctgacttagaatttatctttgatagattttatcgtgtagataaatcacgtgct
cgtagtcaaggaggcaatggattaggactatcaatagcagaaaaattgtgcaacttaac
gggtggtatgattcaagtagaaagtgaactacaaaagtacacgactttcaaatcagtttt
ccagtactaaactaa

5

Sequence 1426

VSYIFSSQITKPIVTMSNMNQIRRDGFQNKLELTNYEETDNLIDTFNEMMYQIEESFN
QQRQFVEDASHELRTPLQIIQGHNLNIQRWGKKDPAVLEESLNISIEEVNRITKLVEELL
LLTKDRVNHNVLECEVDVNSEIQSRVKSLLQHLHPDYTFETHLATKPIQLKINRHQFEQL
LLIFIDNAMKYDTEHKKIKIVTQLKNKMIMIDITDHGMGIPKADLEFIFDRFYRVOKSRA
RSQGGNGLGLSIAEKIVQLNGGMIQVESELQKYTTFKISFPVLN*

10

Sequence 1427

Contig_0604_pos_824_1960,

15 is similar to (with p-value 0.0e+00)

>sp:sp|P50840|YPSC_BACSU HYPOTHETICAL 43.5 KD PROTEIN IN COT
D-KDUD INTERGENIC REGION PRECURSOR. >gp:gp|L47838|BACPONAYPP
_15 Bacillus subtilis (clone YAC15-6B) ponA gene, yppBCDEFG
genes, ypqAE genes, yprAB genes, cotD gene, ypsABC genes, rn
aP gene, yptA gene, ypuA gene, kduDI genes, kdgRKAT genes, y
pwA gene, complete cds's. NID: g1146168. >gp:gp|Z99115|BSUB0
012_158 Bacillus subtilis complete genome (section 12 of 21)
: from 2195541 to 2409220. NID: g2634478.

25

atgtttcaattatttagcagtatgtccaatgggattagaagcagttgtagctaaagaaata
caagaatttaggttacgatacacaagtagaaaaacggtcgtatctttttgaaggatgaa
agtgcattgttagatgttaacttatgtgttacgtaccgcagaccgaataaaaaattgtaatg
ggctcagttcaatgctactactttcgacgaattatttgagcaaaacaaagtcattaccatgg
gagaccgtaattgacaaagaaggaattttcccggttcaagggcggagtgtaaaatcaaca
ttatatagtgtagctgactgccaagcaattactaaaaaggctatagtagaacgacttaag
catgctcatcaagaaaaaggaatggctaagtgaacaggcgccaaatatcctgtagaagta
gcaatattaaaagataatgtattattgaccattgatactgctggttctggactcaacaag
cgaggttatcgctattgctcaaggtgaagcaccattaaagaaacattggcagcaagcctc
attcgtttagcaaatgggaatggtaacacacaccttaattgatcctttctgtggttcaggc
accattgctattgaagcatgtcttattgcacaaaatattgcacctggttttaataggat
tttgatctgaacaatggaatatgatgccacctaattttatgacaaatttcgtgatgaa
gctgatcaattggctgactatgataaagacatacaagtatatgcttctgatattgatcca
gaaatgattgaaattgcaaaacgtaatgcagaagaagttggtctcggagatatcatcacg
ttcaatgttaaagatgtgaatactttgtctattgataccgacaagcctgtagcacttggt
ggcaatccacgctatgggtgaaagaattggagatcgtgaagaagttgaagaaatgtatcgt
tatataggtacacttttgaacaacatcctcatttatcgccctacatattaacaagtaac
aaagaatttgataacttagttaatcgaaaagcgactaaaagacgtaagttgttcaacggg
tatattgaatgtactatcagtatgtggtgtaaaaagcaagtaataaaaattaa

30

35

40

Sequence 1428

MFQLLAVCPMGLAENVVAKEIQELGYDTQVENGRIFFEDESIAIVRCNLWLRADRIKIVM
GRFNATTFDELFEQTKSLPWETVIDKEGNFPVQGRSVKSTLYSVPDCQAITKKAIVERLK
HAHQEKGWLSETGAKYPVEVAILKDNVLLTIDTAGSGLNKRGYRIAQGEAPIKETLAASL
IRLANWNGNTPLIDPFCGSGTIAIEACLIAQNIAPGFNRDFVSEQWNMMPPNIYDKFRDE
ADQLADYDKDIQVYASDIDPEMIEIAKRNAEEVGLGDI IQFNVKDVNTLSIDTKPVALV
GNPPYGERIGDREEVEEMYRIIGTLLKQHPHLSAYILTSNKEFEYLVNRKATKRRKLFNG
YIECTYYQYWGGKKQSNKN*

50

Sequence 1429

Contig_0604_pos_2092_2766,

55 putative peptide of unknown function

atgtatcatattttaattagaaaggggcttttaaatatgaagactgttttgattgtaggc
gcaaatggtagagtatctatcgaagcgacaaaaattttcctagagaactcaagattta
gttgattttatttttgaaaaatgcgcacgtatatacctgattacgcctctaatagaattaaa
gtttatgagggagacgctaaaaatattgaggatttagaaagtgttttaacaatgttgat

- gttggttttgcgaagtttatcgggatcacttgataaacaagctgaaactatcgtaaaagcc
atggataacaaaaatgtaagagactgattttttagcagctcctggattttatgatgaa
ctaccagaaccattcaatcaatggaataaagaacaatttggcgaaaaattgaatcgctac
cgcaaagcatctgatattatagaaaattcagatttagattacacaataatcgcgccaggc
5 tggcttacagataaaaaatgaaaatgtatatgagatcacagcaaagaacgaaacattttaa
ggtagtgaagtatcacgtaaaagtgtagcatcttttagcagtacaaattgccaaaaaccca
gaactacactctaagaaaaatattggtgtgaataaacctaatacagaaggtaataaacct
gcttggttcaattag
- 10 Sequence 1430
MYHILIRKGLNMTVLIVGANGRVSIETKIFLENSRFNVDLFLRNAHRI PDYASNRIK
VYEGDAKNIEDLESALNNVDVVFASLSGSLDKQAETIVKAMDNKNVKRLIFVAAPGIYDE
LPEPFNQWNKEQFGEKLNRYRKASDIIENSOLDYTIIRPGWLTDKNENVYEITAKNETFK
GTEVSRKSVASLAVQIAKNPELHSENIGVNKPNTGKPAWFN*
- 15 Sequence 1431
Contig_0604_pos_3143_6580,
is similar to (with p-value 4.0e-81)
>sp:sp|P54159|YPBR_BACSU HYPOTHETICAL 137.4 KD PROTEIN IN BC
20 SA-DEGR INTERGENIC REGION. >gp:gp|L77246|BACYACA_6 Bacillus
subtilis (YAC10-9 clone) DNA region between the serA and kdg
loci. NID: g1256615. >gp:gp|Z99115|BSUB0012_144 Bacillus su
btilis complete genome (section 12 of 21): from 2195541 to 2
409220. NID: g2634478.
- 25 atggcaataaatgaacaattagatactttatataaaattaaaaaaagaattggaaaaatct
aataacagaccttttaattaataaccattaatcaagtaataaaaaaagtttatttaaatcaa
tacaccgcaacattttagtagacatttctcagctggaaaatcaacggttaattaatttgtaa
ttagaacaagataactaccaagttctccagtagcaacaacgagcaatacagctatcggt
tcagtagccaaggaagacgaaatcattgccaatcttacgcaacaacaatatacaaaatta
30 aaaacgtacaacgatgttaacaaatgaatcgacaaaatgtcgatgtagaatcgatagaa
attaattttccatcaaacaaatttaattctaggtttacatttcaagatacaccgggtgtt
gactctaacgttgcaacgcacagtcgagtagcagagcaatttatgtatacaagtaacttg
ttgttttatacagtagactataatcatgtccaatcagcggtgaactttaaatttatgaaa
cgtattaatgaagttggtataccaattatttttagtgaattatcagattgataaacataat
35 gaagaagaaattacatttgaacttttaaatcaagagtcgaaaaatcaatcaaagactgg
gatatacaacttcaagatacttattacgtttcaagtttgatcatccacagaatgaaatt
gacaaactttcaaattttctagtatttatggatcaacatcgatgaatcaacagaagactat
gttaatagaaccattcaattcattaccgacgcacaatacatatacattcaaaatgaaatg
caatctattcttgacacccttcaaattaatgaagaacaattcgagggaagcatatattcaa
40 tttcaacaaaaatcaagaagtcagcgcagaagcacaattgctcaatgactctaataatta
tttaattattttaaacagaagcgtaagatatattagataatgcttatatcatgacgtac
gatatgctgtaattctttacggaattattttagaagcatggcaactgattttaaagtgaat
ggattttttaataaaaagggaagaaaaagaagaacaaatcaaacgacttaatgaggcg
accactcaattgcaagagaaagttaatcaacaagtagcacaaccacttcgtgaagatatg
45 tcatttttaactagattcataataaacatgctgtgaatgaaaaataactaaatcaagaa
tatgacgtcggtccgtcacttatatcagagctatatcaaaactcaaacgagcatttagcaac
acatacgttttaacattttcagatgaagttataaaaagctttgaataaaaaaatagaaaat
gagtcaacaccactatttgaagaagctgtcaatcatgtacaagttaatgaattatcgagt
gatgaaaatgaagataggtatgaatatgatagatacattgaacttaacacattaaaggat
50 tcgcttacatcccacaactacaaacattactatatccatttagacgattcttttagataaa
ttaattggaagaacagagactcattttgaacttaaaacaagaaaattcaactgcttatcat
cgtaaacatgagacacaacatcgtaacgagtttggttacatctaataagatattaagcgt
gcattagatatcggttaaaagatgtaccattatttgatcgactaaacaagatatcaccgat
accattctgagactcgataatacaataacaaaagttggtgtttttggtacatttagcgca
55 ggtaaaagcagcttaataacgctctactaggtgaaaactatttagttagttctcctaata
ccaacaacagcggcaacgacggaattatcatatggtaaagagagtcacaaatcacattaaaa
tcgaaagaacaattactagaggaagttaatcatgtactagaattttatgaaatatcggtt
aacacatttagacgactttattgagagtgatttagataagttaaaattgaaactagaaaag
aaccaacttgcattttatttagtgcaattgagaacatttatgaaatgtacacatctatgtta

gaacattcacttatacacacagtatcgcttgaagaaattaaaaaatggagtgccgaggat
 gagtatgctactttcgtgaaaactgtacaccttaagctaccttttagattggctcaaggg
 aaaatcattattgattcactcggtttacattccaataatcaaagacacacgaatgaaact
 gaacaaaattctcacttcttctgaccttattttgtatgtaacttacttcaatcattcattt
 5 acagataacgcagaaaagcctttatcgaaacatatgaaagatatgaatcaattaaatgaaaat
 caagcttttaaaatgataattaatgctgttgacctagctgaagataaacaggatatacaa
 gctgttgaggattatgttgagatgcactggggcaagttaacatacactctgaaatatat
 agtgtttcaagtcgtcaaagtttaaatgggaataacattggcataaatgaattaaagaaa
 agtatacaatactttgcaaaggttgaaatccagaacaatttttagagcaacaaatgacttat
 10 caattgcaacaaatgaatacgtccttttaaaaacatgattaaagattttcatgatgacaac
 gcaaaaattgtcagcgagacaacataaaattaaatcactataaaaaatcaaaccgggttaaat
 caagagttgattgatacaactgcacaacgtacttttaatgaagtagaagaacaagtatat
 catctaaatgaacgggttaaaactacaacttttagatgagggttaaatctgtatttaatagt
 cagatgacacaaaataacgacttcaatgaggaaaagaaaatttcaactaaaatatattta
 15 gatcaaatcatcaacgcttattcttagaacaatcacttattacagaaaggattaaaaaa
 tattttaattcacaactagaagaacaaatcataccagtcagaaaaagttaaatcagatt
 catgtttattataaatgcaaaatttaattgtggagccatcaatcggttgatagccattactt
 caaattgaacttaattcaatgttgcaatcactaccaaacagtttaactaaacgtaaaaata
 gttaatccaaagtcacaaaaggatattcaagaacacatagctaataaactcttgaatta
 20 ttacaagatgatttgaactcattgcgtcgacaattaaatgattatatccacgagatgact
 caacttgcggaacatcaatttcaaagtgttgagacatccatccaacaacaaattgatgag
 ttactttcattcactatcgatgatacattaatacaacaactagaattgaaaccacacaa
 ttgataaatatttttatag

25 Sequence 1432

MAINEQLDLYKLKKELEKSNNRPLINTINQVIKKVYLNQYTATFVGHFSAGKSTLINLL
 LEQDILPSSPVPTTSNTAIVSVAKEDEIIANLTQQOYTKLKYNDVKQMNQRQNDVESIE
 INFPSNKENLGFTEQDTPGVDSNVATHQSSTEQFMYSNLLFYTVDYNHVQSALNFKFMK
 RINEVGPIIIFVINQIDKHNEEEITFETFKSRVEKSIKDWDIKLQDTYYVSKFDHPQNEI
 30 DKLSNLFVFMDOHRESTEDYVNRITQIFITDAQYIYIQNEMQSILDTLQINEEQFEEAYIQ
 FQONQEVSAEAQLLNDNSQFLNYLKQKRKDLNAYIMTYDMRESLRNYLESMATDFKVN
 GFFNKRKKKEEQIKRLNEATTQLQEKVNQQVRQPLREDMSFLTRFINKHAVNEKILNQE
 YDVVPSLISELYQTQTSISNTYVLTFSDEVIKALNKKIENESTPLFEEAVNHVQVNELSS
 DENEDRYEYDRIELNLTSLTSHNYKHYYIHLDDSLDKLIGRTETHFELKQENSTAYH
 35 RKHETQHRNEFVTSNQDIKRALDIVKDVPLFDRTKQDITDTILRLDNQITKGVFGTFSA
 GKSSLINALLGENYLVSSPNPTTAATTELSYGKESQITLSKEQLLEEVNHVLEFYEISF
 NTLDDFIESDLKLLKLEKNQLAFISAIEKHYEMYTSMLEHSLIHTVSLEEIKKWSAED
 EYATFVKTVHLKPLDWLKGIIDSLGLHSNNQRHTNETEQILTSSDLILYVTFYFNHSF
 TDNDKAFIEHMKDMNQLNENQAFKMIINAVDLAEDKQDIQAVEDYVADALGQVNIHSEIY
 40 SVSSRQSLNGNNGINELRESIQYFAKVESRTILEQQMTYQLQQMNTSFKNMKDFHDDN
 AKLSARQHKLNHYKNQTRLNQELIDTTAQRTEFNEVEEQVYHLNERLKLQLLDEVKSVFNS
 QMTQNNDFNEEKKISTKIYLDQIHQRLFLEQSLITERIKKYFNSQLEEQIIPVMKKLNQI
 HVIINAKFNVEPSIVDTPLLQIELNSMLQSLPKQLTKRKIVNPKSQKDIQEHIANQTLEL
 LQDDLNSLRRLNDYIHEMTQLAEHQFQMLETSIQQIDELLSFTIDDTLIQQLKTTQ
 45 LDNIL*

Sequence 1433

Contig_0604_pos_6599_7477,

is similar to (with p-value 2.0e-48)

50 >gp:gp|U78771|LLU78771_2 Lactococcus lactis DNA polymerase I
 (polA) gene, complete cds. NID: g2281292.
 atgcctcaaagagtattgctcggttggaatggctttattatttagacatttttatgca
 acaagcttacacaatcaatttatgtacaattctaaaggaattcctacaaacggcattcaa
 ggttttgtaagacatattttttagcgctatcaaagaaatcgaaaccactcacgtagcagtt
 55 tgttgggatatggggcaagaacattcagaaatgaaatgtatgatggttacaacaaaaat
 cgcccagcaccactgatgaacttatacctcaatttgattatggttaaagaatatcgcat
 cagtttggtttgtaaatgttggaagcgtaattatgaagctgatgatattattggtagt
 ttagcggaacatatattcacaagaacatgaagtttatatcattaccggagaccgagactta
 ctcaatgcattaatcataatgtagaagtttggcttattaaaaaagggtttacaatatat

caacggttacacgcttgatcggttttattgatgaatatggactaaatcctcaacaattaata
 gatgtcaaagcttttatgggtgatactgcagatggctattctgggtgtaaaagggataggt
 gaaaaaacagcaattaaattaattcaaaatcatggactgtcgaaaatgtagtgaacaat
 ttatcatcattaactcccgctcaacagaaaaataacaaataatttaaatcatctgcat
 5 ttatcaaaatcactcgagaaatatataccaaagttccaattgaaacagacaaactttt
 aaagagatgacatatgctcatacactaaatgagattttatccatttgtaatgaacatgaa
 ctatacggttcaagtaaatatattgcaactcacctctaa

Sequence 1434

10 MPQRVLLVDGMALLFRHFYATSLHNQFMYSKGIPTNGIQGFVRHIFSAIKEIEPTHVAV
 CWDMGQETFRNEMYDGYKQNRPAAPPDELIPQFDYVKEISHQFGFVNVGKRNIEADDIIGS
 LAETYSQEHEVYIITGDRDLLQCINHNVEVWLIKKGFTIYQRYTLDRFIDEYGLNPQQLI
 DVKAFMGDTADGYSVGKIGEKTAIKLIQNHGTVENNVNLSLTPAQKKITNNLNHLH
 LSKSLAEIYTKVPIETDKLFKEMTYAHTLNEILSICNEHELYVSSKYIATHL*

Sequence 1435

Contig_0604_pos_7659_0,

putative peptide of unknown function

atgaagagcaaacccgaaattaaacggctcggaacatctgctcttttttattgagcaaatgt
 20 atgagttattcattgtcaaaattatcaacattaaaaacgtataattttcaaatacacatca
 aacaacaaagaaaaaacatcaagaataggagtggcaatagctttgaataatcgtgataaa
 ttacaaaaatttagtattcgaaaatacgcaattggaacattttctactgtgattgcaaca
 cttgtgttcatgggtatcaatacaaacatgcaagtgcgcgacgagttgaatcaaaatcaa
 aagtttaattaaacaattaaatcaaacagatgatgatgattcgaatacgcatagtcaagaa
 25 atcgaaaaatacaaaacaaaattcttagtggcgagactgaatcattacgttcatcaactagt
 caaaatcaagcaaatgcacgactgtcggtatcaattcaaagacactaatgaaacatcgcaa
 caattacctacaaatgtttcggtatgatagatcaatcaatcgcatagtgaagcaaatatg
 aataacgaaccattgaaagttgataatagtactatgcaagcacatagtaaaatagtaagc
 gatagcgatgggaatgcttctgaaaaataaacatcataaactaacagaaaatgtacttgca
 30 gaaagccgagcaagtaaaaaatgacaaagagaaagagaatctacaagagaaagataaatcg
 cagcaagtacatccaccattagataaaaaatgcattacaagcttttttgacgcacatcat
 cacaattacagaatgattgatagagatcggtcggtatgcaacagaatatcaaaaagtcaaa
 tctacttttgactacgtcaatgacttactaggttaataatcaaaatattccttcagaacag
 cttgtttcggcatatcaacaattagagaaagcattagaacttgacgtagcttaccacaa
 35 cgatctactacagaaaaacgtggtagaagaagtagcagaagtggtgttgagaatcggtca
 tcaagaagcgattacttagatgctagaactgaatattatgtttcaaaagacgatgatgat
 tctggtttccctcctggtactttcttccatgcttcaaatagaagatggccttataattta
 ccaagatctaggaacatcttacgtgcttctgatgtacaaggtaatgcttatatcactaca
 aaacgacttaagatggatataaatgggatattttatttaatagtaaatcataaagggcac
 40 gaatatatgtactattggtttggacttccaagtgatcaaacaccaactgggtccagtaact
 ttcactattatcaaccgtgatggttcaagtacatctactggtggcggttgatttggtatca
 ggtgcaccactacctcaattttggagatcagcaggtgctatttaattctagcgtagcgaat
 gatttttaaacatggctccgctacaaattatgcattttatgatggtgttaataattttct
 gactttgctagagggggagaattatacttcgacagagaaggcgctacacaaactaataaa
 45 tattatggcgatgaaaacttcgcattgctaaatagtgaagaaaccagatcaaaataagagga
 ttagatacaatatatagttttaaaggtagtggtgatgtaagttatcgattttcattttaa
 actcaaggagctccaactgcaagattgtattatgctgctggcgcgcttctggtgaatat
 aaacaagcaacgaactataaccaactctatgtcgaacctataagaattatcgaaatcga
 gtacagtcgaatgtccaagttaaaaatcgtacacttcattttaaagaacaatcagacaa
 50 ttcgatcctacattacagagaactactgatgttccatttttggtatgtgacggttccgga
 agtattgatt

Sequence 1436

55 MKSKPKLNGRNICFLLSKMSYSLSKLSTLKTYNFQITSNNKEKTSRIGVAIALNNRDK
 LQKFSIRKYAIGTFSTVIATLVFMGINTNHASADELNQNKLIKQLNQTDDDDSNTHSQE
 IENNKQNSSGQTESLRSSTSQNQANARLSDQFKDNETSQQLPTNVSDDSINQSHSEANM
 NNEPLKVDNSTMQAHSKIVSDSDGNASENKHKLKTENVLAESRASKNDEKENLQEKDKS
 QQVHPPLDKNALQAFDASYHNYRMIDRDRADATEYQKVKSTFDYVNDLLGNNQNIPEQ
 LVSAYQQLEKALELARTLPQRSTTEKRRSTRSVVENRSSRSDYLDARTEYYVSKDDDD

SGFPPTFFHASNRRWPYNLPRSRNLRASDVQGNAYITTKRLKDGQWDILFNSNHKGH
 EYMYWFGPLPSDQPTPTGPVFTTIINRDGSSSTSTGGVGFSGAPLPQFWRSGAINSSVAN
 DFKHGSATNYAFYDGVNNFSDFARGGELYFDREGATQTNKYGDENFALLNSEKPDQIRG
 5 LDTIYSFKGSGDVSYRISFKTQGAPTARLYYAAGARSGEYKQATNYNQLYVEPYKNYRNR
 VQSNVQVKNRTLHLKRTIRQFDPTLQRTTDPILDSGSGSIDX

Sequence 1437

Contig_0604_pos_1816_1418,

is similar to (with p-value 4.0e-34)

10 >sp:sp|P50840|YPSC_BACSU HYPOTHETICAL 43.5 KD PROTEIN IN COT
 D-KDUD INTERGENIC REGION PRECURSOR. >gp:gp|L47838|BACPONAYPP
 _15 Bacillus subtilis (clone YAC15-6B) ponA gene, yppBCDEFG
 genes, ypqAE genes, yprAB genes, cotD gene, ypsABC genes, rn
 aP gene, yptA gene, ypuA gene, kduDI genes, kdgRKAT genes, y
 15 pwA gene, complete cds's. NID: g1146168. >gp:gp|Z99115|BSUB0
 012_158 Bacillus subtilis complete genome (section 12 of 21)
 : from 2195541 to 2409220. NID: g2634478.
 atgaggatgttgtttcaaaagtgtacctatataacgatacatttcttcaacttcttcacg
 atctccaattctttcaccatacgggtgattgccaacaagtgtctacaggcttgtcggatc
 20 aatagacaaagtattcacatctttaacattgaactgtatgatctccgagaccaacttc
 ttctgcattacgttttgcaatttcaatcatttctggatcaatatcagaagcatataacttg
 tatgtctttatcatagtcagccaattgatcagcttcacgaaatttgtcataaatatt
 aggtggcatcatattccattgttcagatacaaaatccctattaaaaccagggtgcaatatt
 ttgtgcaataagacatgcttcaatagcaatggtgcctga

Sequence 1438

MRMLFQKCTYITIHFNFFTISNSFTIRWIANKCYRLVGINRQSIHIFNIELYDISETNF
 FCITECNFNHFWINIRSIYLYVFIIVSQLISFITKFVINIRWHHIPLFRYKIPKTRCNI
 30 LCNKTCFNNGA*

Sequence 1439

Contig_0605_pos_816_1457,

is similar to (with p-value 4.0e-52)

35 >gp:gp|Z99122|BSUB0019_48 Bacillus subtilis complete genome
 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 >gp:gp|U56901|BSU56901_2 Bacillus subtilis putative transcri
 ptional regulator (yvhJ), Ycr59c/YigZ homolog (yvhK), histid
 ine kinase (degS), transcriptional regulator of degradation e
 nzyme (degU), (degV), (comFA), (comFB), (comFC), flagellar p
 40 rotein (yviB), negative regulator of flagellin (flgM), flage
 llar protein (yviC), flagellar-hook associated protein 1 (fl
 gK), flagellar-hook associated protein 3 (flgL), (yviE), tra
 nsmembrane protein (yviF), (csrA), flagellin (hag), flagella
 r protein (yviH), flagellar hook-associated protein 2 (fliD)
 45 , flagellar protein (fliS), flagellar protein (fliT), sigma-
 54 modulator homolog (yviI), and (secA) genes, complete cds.
 NID: g1762326.
 atggataaatccataattactatttaaacagcacattcaattgaaaatgtgataagtaaa
 tcacgctttatagcatatattaagcctgtttcgactgaaaatgaagcaaaagctttata
 50 gatgaaattaaaacaaaacataaagatgcaactcataattgttcagcctatactgtcggg
 ccagagatgaatattcaaaaggcaaacgacgatggcgaaccaagtggaaacagctggcatc
 ccaatgcttgaaatactgaaaaaacaagagatacacaaatgtttgtgtcgtcgtgacacgc
 tacttcggtggtatcaagtttaggtgcaggcggcttattagagcatatagcggcgccgtg
 cgtgatgtgatatatgatataggtagagtcgaactaagagaagctattccagtaaccggt
 55 acgttagattatgatcagacaggtaaatttgaatatgaacttgcttactacattctta
 ttaagagaacaattttataaccgataaagtaagttatcaaattgacgtagtaaaaaatgaa
 tatgatgcttttatagacttttttaaatcgaattacttctggaaattatgatttgaacaa
 gaagaccttaaaactattaccttttgatattgaaaccaattaa

Sequence 1440

MDKSIITIKQAHSENVISKSRFIAYIKPVSTENEAKAFIDEIKTKHKDATHNCSAYTVG
PEMNIQKANDDGEPSGTAGIPMLEILKKQEIHNVCVVVTRYFGGIKLGAGGLIRAYSGAV
RDVIYDIGRVELREAI PVTVTLDYDQTGKFEYELASTTFLLRQFYTDKVSQIDVVKNE
5 YDAFIDFLNRITSGNYDLKQEDLKLPPFDIETN*

Sequence 1441

Contig_0605_pos_3202_3879,
putative peptide of unknown function

10 gtgggcatttttagtatacggggtcagggatagcgagtgtaacaaacaataactcacgca
aaagaaagtcacgattcaactcctcaaaatattaaattagtggaacgtatgatacttct
caagttgattccaaaacgatgaaacaatttaagaaatagaaaaagaagataataatttc
cacataactaaacatggaaataaagtcggtttagaagacaaattacctaataccagagaat
aaaacttcaagttattcagctgatggtagtgctgaaaataatacaaaagtaattaatttc
15 tctgattttgttggaatatggatgggaaagatgatggaaaaatcggatgggataacc
ttttatagtggttaaatcatataacggacaacacgatgggtcaaaaagtaaaaaagggact
catgtacattgtaatagatttaacggaacaaaatctgatcatagatactgggtcaaaaaa
catcctagagcttattgtagatttttataaaagtgattgctggtatcacgccaaagcttat
aaatgttcttcttgggaaaaatgactaaatgcgatggtttgaatagattttatagaaaa
20 ggtgtcaaaagattgctcatcatggaaaggtaaacccaaacataaaaaactggcctaaaaa
gcatggtatagaaattaa

Sequence 1442

25 VGILVSGSGIASVQTNITHAKESHDPQNIKLVGTYDTSQVDSKTMKQFKEIEKEDNNF
HITKHGNKVVEDKLPNPENKTSSYSADGSAENNTKVINFSDVFGNMDGKDDGKISDGIT
FYSGKSYNGQHDGQKVKKGTHVHCNRFNGTKSDHRYWSKKHPRAYVDFYKSDCWYHAKAY
KCSSLGKMTKCDGLNSIYRKGVKDCSSWKGKPKHKNWPKTAWYRN*

Sequence 1443

30 Contig_0605_pos_3883_4245,
putative peptide of unknown function
atgaataaaatcttaaaaatattaacttctattattgttatcattattaccttaaca
gtttggacttttagtggtattacttatcagaacacaaagagtgagaaaatcatcaatcac
gttatagaacgtaagggttgggataaaaaataaaaaatgaaaaatgagttttaattatt
35 ataatgggatattgctgaaaaagatattgtttttaagatcaaccatatagtgagtatgag
tataacgtgacaccagcaccatggacagatgataaagaatataagggtgtggggggaaca
gatttacaaaagaaagactcctattataaatatcttttagaatcagaaccttacagaaaa
taa

40 Sequence 1444
MNKILKILITSIIIVIIITLVWTFVITYQKHKSEKIINHVIERKGWDKKIKNEKMSFNI
IMGYAEKDIVFKDQPYSEYENVTAPWTDKEYKVWGETDLQKKDSYKYKLESEPYRK
*

Sequence 1445

45 Contig_0605_pos_4375_4740,
putative peptide of unknown function
atgaaatgggttaaaaatcattatattaattttatcttttagttccaatagaattttatcgga
ttatttactgattatcaaacagggtttattaataggttacatcccattttatagtagttgct
50 atattaatgagtatatccatttttaaaatttggttttaaaaaataataaagtattgtaata
ggtagatgcattggaatatttcttcttggggatgtgttcatttatttatgaacatttac
gattcatcagattattttaaacctcttaccacagatattttcgcattatttttaggagct
attcattcatagttattatgcttatatatatttagttatctatggtttttctcatcgcaat
aattaa

55

Sequence 1446

MKWLKIIILILSLVPIEFIGLFTDYQTGLLIGYIPFIVVAILMSISIFKFGFKNNISIVI
GRCIGIFLSWGCVHFLFMNIYDSSDYFKPLTTDIFALFLGAIHFIVIMLIYLYVYGFSHRN
N*

Sequence 1447

Contig_0605_pos_5855_5520,
is similar to (with p-value 3.0e-38)

- 5 >gp:gp|AF051916|AF051916_2 Staphylococcus aureus plasmid pJE
1 remnant of replication protein Rep (rep), trimethoprim res
istance protein DfrA (dfrA), thymidylate synthetase ThyE (th
yE), and putative transposase Tnp (tnp) genes, complete cds;
and unknown gene. NID: g3676404.
- 10 atgaatgatataactaaacgtttattaaaaccaataattaatgagctttcttcaattttt
aataaccttcattattaataagatcaaagctaaaaaggacgtaaaattgaatggttagag
tttacctttgacgctgagaaacgcattcacacaagcgacaaccacaaatgactaatata
ggtaagtcgagccaatataccaatcgtagaaaaaacctaaatggttagacgaaaagata
tataaacaatctcaagagatacataatgaagatgcaaaattaaaacaagatcgagaggca
15 tttcaacgtcaattagaagaaaaatgggaggaataa

Sequence 1448

MNDITKRLKPIINELSSIFNNLHINKIKAKKGRKIEWLEFTDAEKRIHNKRQPMQNTNI
GKSRQYTNREKTPKWLDEKIYKQSQEIHNEDAKLKQDREAFQRQLEEKWEE*

20

Sequence 1449

Contig_0605_pos_2569_1520,
is similar to (with p-value 0.0e+00)

- 25 >pir:pir|A55856|A55856_1lm protein - Staphylococcus aureus
atgatagctcagtttaataattacaccattattattgtaatatcaaaaaattagattta
gtagatcgctcctaatttcagaaaagtagacatacgaaacctatctcagtgatgggaggaacg
gtcattttattttctttcttaataagggatttggctcggacaccctattgaacgtgaggtt
aaaccgcttatattaggtgcaattacaatgtatatggttggttgattgattgatattttac
gatctaagacctattttaaggttagcaggtcaaattggttcagctttaattggttacgttt
30 tatggaattacaatagactttatttcattgccaattggtccaacgattcattttggcata
ttcagcattcctattacagtaatatggattgtagcaattaccaatgctattaatcttate
gacggacttgacttgccctcaggcgtctcagcattggcattaatgactattggattc
atcgctattttacaagcgaacatatatttatcatgatttgcgtgtgacttttagggctc
ttacttggtttcttattctataactttcaccacgcgaaaattttcctaggtgatagtggt
35 gcattaatgataggatttattatcggtttcttatecttactcggctttaagaatatcaca
tttattgcattattctttcctatagttatattagcgggtgccattttattgatacattattt
gcaatgattcgtcgaatgaaaaaagggcaacatataatgcaagcgggacaagtcacattta
catcataaattacttgccttaggatatacgcatagacaaaccgttttacttatttattca
atagcgattatgttttagtttatctagtggtatcctctatttatcccaaccgttgggtgca
40 cttatgatgttcattctcattgtctttacgattgagttgatcggtgaatttactggatta
atagatgataattatcgaccaatattaaatttaattacaaaaaaggaaatggttaagcaa
catcattatgatgagcatcaccgttcataa

Sequence 1450

- 45 MIVSLIITPIIIVISKKLDLVDRPNFRKVHTKPIISVMGGTVILFSFLIGIWLGHPIEREV
KPLILGAITMYMVGGLIDDIYDLRPYLKLAGQIVAALIVTFYGITIDFISLPIGPTIHFGI
FSIPITVIWIVAITNAINLIDGLDGLASGVSAALMTIGFIAILQANIFIIMICCVLLGS
LLGFLFYNFHPAKIFLGDGALMIGFIIGFLSLLGFKNITFIALFFPIVILAVPFIDTLF
AMIRRMKKGQHIMQADKSHLHKLLALGYTHRQTVLLIYSIAIMFSLSSVILYLSQPLGA
50 LMMFILIVFTIELIVEFTGLIDDNYRPILNLITKKGNGKQHHYDEHHS*

Sequence 1451

Contig_0605_pos_0_673,
is similar to (with p-value 2.0e-43)

- 55 >sp:sp|P32436|DEGV_BACSU DEGV PROTEIN. >pir:pir|S28596|D3019
1 hypothetical protein U3 - Bacillus subtilis >gp:gp|Z18629|
BSCOMFG_1 B.subtilis comF gene. NID: g39847. >gp:gp|Z99122|B
SUB0019_45 Bacillus subtilis complete genome (section 19 of
21): from 3597091 to 3809700. NID: g2636029. >gp:gp|U56901|B

- SU56901_5 *Bacillus subtilis* putative transcriptional regulator (yvhJ), Ycr59c/YigZ homolog (yvhK), histidine kinase (degS), transcriptional regulator of degradation enzyme (degU), (degV), (comFA), (comFB), (comFC), flagellar protein (yviB), negative regulator of flagellin (flgM), flagellar protein (yviC), flagellar-hook associated protein 1 (flgK), flagellar-hook associated protein 3 (flgL), (yviE), transmembrane protein (yviF), (csrA), flagellin (hag), flagellar protein (yviH), flagellar hook-associated protein 2 (fliD), flagellar protein (fliS), flagellar protein (fliT), sigma-54 modulator homolog (yviI), and (secA) genes, complete cds. NID: g1762326. atgaagattgcagttatgaccgattctacaagttattaccacaacatataatagaacaa tataacataccagtcgcttcactaagtgttaactttcgtatgatggagtgaaatttcactgag agtgatgatttttctgtagatgatttttataaaaaaatggcttcatctaaaactatacca acaacaagccaacctgctattggcgattggattgaaaattttgagagattaagagaacaa ggatacactgatgtcatcgtagtaacttatcaagtggataagcggagctatccttca gcaacacaagctgggtgaaatgggtgaagatattcaagtacatacgtttgatagccgtctt gctgcgatgattgaaggtagctttgcaatttacgtgctcaattgggtacaaaaggatata aaactgatgatattattaatgaactaactgaaataagacaacatattgggtgcatactta attgttgatgatttaaaaaatttacaaaaaagtggtcgatcactggagctcaagcttgg gtaggtacattattgaaaatgaaacctgtcttgcgttttgaagaagatggtaaaatcat ccacacgaaaagtagctactaaaaaacgtgcgctaaaatcttagaaacaaacattttt aaagaaatagaGT
- 25 Sequence 1452
MKIAVMTDSTSYLPQHIEQYNIPVASLSVTFDDGVNFTESDDFSVDDFYKKMASSKTIP
TTSQPAIGDWIENFERLREQGYTDVIVINLSSGISGSPSATQAGEMVEDIQVHTFDSRL
AAMIEGSFAIYAAQLVQKGYKPDDIINELTEIRQHIGAYLIVDDLKNLQKSGRITGAQAW
VGTLLKMKPVLRFEEGDKIHPHEKVTRTKKRALKSLETNIFKEIEX
- 30 Sequence 1453
Contig_0607_pos_339_716,
is similar to (with p-value 2.0e-41)
>sp:sp|P18158|GLPD_BACSU AEROBIC GLYCEROL-3-PHOSPHATE DEHYDR
35 OGENASE (EC 1.1.99.5). >pir:pir|C45868|C45868 glycerol-3-pho
sphate dehydrogenase (EC 1.1.99.5) - *Bacillus subtilis* >gp:g
p|M34393|BACGLPKD_3 *B. subtilis* glycerol kinase (glpK) and gl
lycerol-3-phosphate dehydrogenase (glpD) genes, complete cds.
NID: g142990. >gp:gp|Z99108|BSUB0005_198 *Bacillus subtilis*
40 complete genome (section 5 of 21): from 802821 to 1011250. N
ID: g2633055. >gp:gp|Z99109|BSUB0006_5 *Bacillus subtilis* com
plete genome (section 6 of 21): from 999501 to 1209940. NID:
g2633260. >gp:gp|Y14079|BSY14079_5 *Bacillus subtilis* chromo
somal DNA, region 75 degrees: glpPFKD operon and downstream.
45 NID: g2226133.
atgtcattatctacattgaaaaggatcatattaaaaagaatttaagagacactgaatac
gatgttgttatcgtaggtggcggtattacaggtgcaggtattgcttttagatgcaagtaac
cgtgggatgaaggtagcttttagtagagatgcaagactttgcacaaggtacaagttcacgc
tcaactaaacttgtaacacggtgggttaagatatttaaaacaactgcaagtaggggtagtt
50 gcagaaacaggtaaagaacgtgctattgtttatgaaaatggccacatgtgacaacacca
gaatggatgcttttacctatgcataaagggtgtacatttggtaaattctcaacttctatt
ggactagctatgtactga
- 55 Sequence 1454
MSLSTLKRDIKKNLRDTEYDVVIVGGGITGAGIALDASNRGMKVALVEMQDFAQGTSSR
STKL VHGLRLYLKQLQGVVAETGKERAIYENGPHVTTPEWMLLPMHKGSTFGKFSTSI
GLAMY*
- Sequence 1455

Contig_0607_pos_2046_3350,
is similar to (with p-value 0.0e+00)
>sp:sp|P18158|GLPD_BACSU AEROBIC GLYCEROL-3-PHOSPHATE DEHYDR
OGENASE (EC 1.1.99.5). >pir:pir|C45868|C45868 glycerol-3-pho
5 sphate dehydrogenase (EC 1.1.99.5) - Bacillus subtilis >gp:g
p|M34393|BACGLPKD_3 B.subtilis glycerol kinase (glpK) and gl
ycerol-3-phosphate dehydrogenase (glpD) genes, complete cds.
NID: g142990. >gp:gp|Z99108|BSUB0005_198 Bacillus subtilis
complete genome (section 5 of 21): from 802821 to 1011250. N
10 ID: g2633055. >gp:gp|Z99109|BSUB0006_5 Bacillus subtilis com
plete genome (section 6 of 21): from 999501 to 1209940. NID:
g2633260. >gp:gp|Y14079|BSY14079_5 Bacillus subtilis chromo
somal DNA, region 75 degrees: glpPFD operon and downstream.
NID: g2226133.
15 atgtacgatcgtctagctggtgtcaaaaaatccgaacgtaaaaaatgttatctaagcaa
gaaacgttaaataaagaaccttttagttaaacgtgatgattaaaaggcggtggctactat
gtggaataccgcactgatgatgcgcgtttaactattgaagtattgaaaaaagctgctgaa
aatggagcagaaatcattaattatacaaaatcagaacacttcacttatgattccaataag
aaagtaaattggtattgaagtattggatatgattgatggcgaaacgtatgcgattaaagct
20 aaaaaagttattaatgcttctggtccttgggttgatgaagtgagaagtgcgattatgca
cgtaacaataagcaattaagattaactaaaggtgtacacgttggttatagatcaatctaaa
ttcccataggtcaagcagtttactttgatactgaaaaagacggacgcatgatttttgcg
attccacgtgaaggaaaaagcttatgttaggaacaactgacacgttttatgataatgaaaa
gcaacacctttaacaacacaagaagatagagactacttaattaatgcaattaactatatg
25 ttcccaacagttaatgttaagatgaagatattgaatcaacatgggctggtattcgctcg
ctaattcttgaaaaaggtaaagatccttctgaaatctcacgtaaagatgaagtttgggaa
ggtgaatctggattattaactatagcaggcggtaaatgaactgggttatcgatcatatggca
ctagaaattggtgatttattagctaaacgttttaaaacaagaatacggattgaaatttgaa
tcatgtgccacaaaaaatctaaaaatttccgggtggtgacgttggcggaagcaaaaacttt
30 gaacactttgttgaacaaaaagttgatgcagctaaaggatttgggaattgatgaagatgtg
gcacgtcgctttaacagtaaatatggttcaaatgttgatcaactatttaatttgctcaa
acggcaccatatacatgatagtaaattaccattagaaatttatgttgaattagtttatagt
attcaacaagaaatggtttacaacaaactgacttcttagtacgtcgttctggcaatta
tactttaattattcaagatgtggttagattataaaaatgctgtgatagatgttatggcggat
35 atgcttaattatagtgaactcaaaaagaagcttatactgaagaagtagaagttgcgatt
gatgaggcacgtacaggtaatgatcaacctgcaactaaagcttaa

Sequence 1456

MYDRLAGVKKSERKKMLSKQETLNKEPLVKRDGLKGGGYVEYRTDDARLTIEVMKKAAE
40 NGAEIINYTKSEHFTYDSNKKVNGIEVLDMIDGETYAIKAKKVINASGPWVDEVRSOYA
RNNKQLRLTKGVHVVIDQSKFPLGQAVYFDTEKDGRMIFAIIPREGKAYVGTDTDFYDNEK
ATPLTTQEDRDYILINAINYMFPTVNVKDEDIESTWAGIRPLILEKKGDPSEISRKDEVWE
GESGLLTIAGGKLTGYRHMALEIVDLLAKRLKQEYGLKFESCATKNLKSIGGDVGGSKNF
EHFEVQKVDAAKGFGIDEDVARRLASKYGSNVDQLFNIAQTAPYHDSKLPLEIYVELVYS
45 IQQEMVYKPTDFLVRSGKLYFNIQVDLDYKNAVIDVMADMLNYSETQKEAYTEEVEVAI
DEARTGNDQPATKA*

Sequence 1457

Contig_0607_pos_3625_4500,
50 putative peptide of unknown function
gtgaaaattgataaaagcaaaaaagagtactattggtattgttcatcttttccatggcatg
gctgaacatatggaccgttatcaagagtttagttgaggctttaatacacaaaggttatgac
gttgtagacataatcacctggacatggtaaagaaatagatgagaatgaacgtggtcat
tttaatagcatgaatcaaattgtagatgatgcttatgaaattattgagacattatatctt
55 gaagagctcaatgtaccctatattatcataggtcattcaatgggctccattattgctaga
tcatttgttgaaaagtatcctgacattgctcaaggtttaattcttacaggaacaggtatg
ttccccaagtggaaaggtgtaccaatacgttttagcaatgaagttagttacattttttt
gggaaacgacgtcgactcaagtggtgaatcaattatgaataaaactttcaataaaaaa
atcactcaacctcgaacagatagtgattggtatttctacacgtcaggatgaagttgataaa

tttgtggaagatgaattttgtggattcaaagtatctaatacagctcatTTtatcaaacttta
 aagaccatgatgaagacagtagaacgacaacaactaaaaagaatggacaaagaactacct
 atactatttatttctgggaaagatgatccttttgggtgaatatggtaaaggtataaaagcat
 ttagctagattatataaaaagcaggtattaaacatataacagtacaactatataaacat
 5 aagcgtcatgaaatattatttgaagaagattatttgaaaacatggcaacacatgtttgaa
 tggatggaaaagcaaattttgaaaaaacaagtgga

Sequence 1458

VKIDKAKKSTIGIVHLFHGMAEHMDRYQELVEALNTQGYDVVRHNRHGKEIDENERGH
 10 FNSMNQIVDDAYEIIETLYLEELNVPYIIIGHSMGSI IARSFVEKYPDIAQGLILTGTGM
 FPKWKGVPIRLAMKLVTFIFGKRRRLKWNQLLNKTTFNKKITQPRDSDWISTRQDEVK
 FVEDEFCGFKVSNQLIYQTLKTMKTVERQQLKRMDELPILFISGKDDPFGEYGKGIKH
 LARLYKRAGIKHITVQLYKHKRHEILFEEDYLKTWQHMFWEKQILKKQK*

15 Sequence 1459

Contig_0607_pos_4521_5510,
 is similar to (with p-value 3.0e-43)
 >gp:gp|AL022268|SC4H2_12 Streptomyces coelicolor cosmid 4H2.
 NID: g3036873.

20 atgacaaagccttttttaaatcgttattgttaggtccaactgcttcagggtaaaactgagtt
 agtattgaagttgctaaaaaatttaattggagaaattattagcggagattcaatgcagggtc
 tatcaaggaatggatattggtacagcaaaagttacaactgaagaaatggaaggtatacca
 cattatatgatagatattttgcctccagatgcttccttttctgcatatgaatttaaaaaa
 agggcagaaaaatatattaaagatattactagaagaggcaaggtgcctattatagcagga
 25 ggaacaggactatatatacaatctctcttatacaactatgcttttgaagatgaatccata
 tctgaagataaaatgaaacaagttaaattaaagttaaaagaacttgagcatctaaataat
 aataagctccacgaatatattagcttcattcgacaaagaatcagccaaggatatacatcct
 aataacagaaaaagagtggttgcgagcaatagaatattattgaaaaacaaaaaactttta
 agttctcgcaagaaagtgcacaatttactgaaaattatgatacattaltaatagggatt
 30 gaaatgtcgcgtgaaacattatatttaagaataaaataaacgtgttgatattatgttgggc
 cacggattatttaattgaagtgaacattctcgttgaacaaggttttgaagcagtgcaaggt
 atgcaagccattgggtataaaagagcttgaccggttattaagggaaatataagcatggaa
 aatgctgtgagagaaattaaaacagcattctcgacaatatgctaaaagacagttgacttgg
 tttaaaaataaaatgaatgttcattggttaaataaagaaaggatgtcacttcaaatgatg
 35 ttagatgagattacaacccaataaaataaaaggagttctaaccatgattgcaaacgaaaa
 catccaagaccaagcactagagaacttta

Sequence 1460

MTKPFLIVIVGPTASGKTELSIEVAKKFNGEIIISGDSMQVYQGM DIGTAKVTTEEMEGIP
 40 HYMIDILPPDASFSAFYEKKRAEKYIKDITRRGKVPIIAGGTGLYIQSLLYNYAFEDESI
 SEDKMKQVKLKLKELEHLNKNKLHEYLASFDESAKDIHPNNRKRVLRAIEYYLTKKKLL
 SSRKKVQQFTENYDTLLIGIEMSRETLYLRINKRVDIMLGHGLFNEVQHLVEQGFASQS
 MQAIGYKELVPVIKGNISMENAVEKLKQHSRQYAKRQLTWFKNKMNVHVLNKERMSLQMM
 LDEITTQINKRSSNHDCRKHPRPSTREL*

45

Sequence 1461

Contig_0607_pos_6502_7752,
 is similar to (with p-value 3.0e-85)
 >gp:gp|U66480|BSU66480_2 Bacillus subtilis SpoVK (spoVK), Yn
 50 bA (ynbA), YnbB (ynbB), GlnR (glnR), glutamine synthetase (g
 lnA), YnaA (ynaA), YnaB (ynaB), YnaC (ynaC), YnaD (ynaD), Yn
 aE (ynaE), YnaF (ynaF), YnaG (ynaG), YnaH (ynaH), YnaI (ynaI
), YnaJ (ynaJ), xylan beta-1,4-xylosidase (xynB), xylose rep
 repressor (xylR), xylose isomerase (xylA), xylulose kinase (xyl
 55 B), YncB (yncB), YncC (yncC), YncD (yncD) and YncE (yncE) ge
 nes, complete cds. NID: g1750106. >gp:gp|Z99113|BSUB0010_37
 Bacillus subtilis complete genome (section 10 of 21): from 1
 781201 to 2014980. NID: g2634090.
 atgtataacgtgacacagcatgacgacttatagaacaaaaataaacgagaaactgctgta

ttaatcgggtgtacatgctcaaacggatcgtcaatttaattttgaatctactatggaagag
 ctcgatgctttatcacaaacttgccaacttaatgttaaaggacaaatcactcaaaataga
 gagcaatttgaccataaatattatgttggaaggaaaaatcgatgaaataaaatctttc
 atagaattccatgatatagatgttgctgtaaccaacgatgaattaacgacggcacagtct
 5 aaacggttaaatgataatttgggcattaaaaatcatcgatagaaccaatatttttagag
 atattcgcgttgcgagcgagaagttagagagggaagctacaagtagaacttgacaaactc
 gattatttgttaccagactacatgggtcatggtaaaagcctatctcgtcttggtggtggc
 ataggaacaagaggcccaggtgaaacaaaattagaatggatcgtcgccatattagaaca
 cgtatgaatgagattaaacatcaattaaaaacgggtcgtggatcatcgggaaagatataga
 10 aataaacgtgaacaaaatcaagtttttcaaatcgcttttagttggttatatacaaatgcagga
 aaatcgctcatgggttaattgttttagctaatgaggagacctatgaaaaaaatattttggtt
 gcaacattagatcccaaaacacgacaaatacaagtgaatgaaggatttaatttaattatt
 tctgatacggtaggattttattcagaaattaccaacgacattggtggctgcgtttaaatct
 acactagaagaagctaaagggtgcagacgtacttatgcatgtcgtcgatgcaagtcattcg
 15 gaataccgtactcaaattgacactgtaaatcaaattattaatgatttagatatggacat
 attccacaagtagtatttttaataaaaaagacttatgtaacgaacagatggatgtacct
 gtatctaaatctgcgcgtgtttttgtatctagtcgtgatgaaatgataaacaagggtg
 aaaaatttagtaattcaagaaataaaaaatagtcctcagccatacgaagaaattgtagat
 agtgctgatgcagatagattatattttcttaacaacacacgcttggtactgaattaata
 20 tttagcgaacacaagcatcttatcgatcaaggatttaaaaaattataa

Sequence 1462

MYNVTQHATYRTKNKRETAFLIGVHAQTDRQNFESTMEELDALSQTCQLNVKGQITQNR
 EQFDHKYYVGKGKIDEIKSFIEFHDIDVVVTNDELTTAQSKTLNDNLGIKIIDRTQLILE
 25 IFALRARSREGKLQVELAQLDYLLPRLHGHGKSLRLGGGIGTRGPGETKLEMDRRHIRT
 RMNEIKHQKLTVDHRRERYRNKREONQVFQIALVGYNAGKSSWFNVLANEETYEKNILF
 ATLDPKTRQIQVNEGFNLIISDTVGFQIKLPTTLVAAFSTLEEAKGADVLHVVDASHS
 EYRTQIDTVNQIINDLMDHI PQVVI FNKKDL CNEQMDVPVSKSAHV FVSSRDENDKQKV
 KNLV IQEIKNSLSPYEEIVDSADADRLYFLKQHTLVTELI FDETQASYRIKGFKKL*

Sequence 1463

Contig_0607_pos_7768_8277,
 is similar to (with p-value 3.0e-34)
 >gp:gp|U66480|BSU66480_3 Bacillus subtilis SpoVK (spoVK), Yn
 35 bA (ynbA), YnbB (ynbB), GlnR (glnR), glutamine synthetase (g
 lnA), YnaA (ynaA), YnaB (ynaB), YnaC (ynaC), YnaD (ynaD), Yn
 aE (ynaE), YnaF (ynaF), YnaG (ynaG), YnaH (ynaH), YnaI (ynaI
), YnaJ (ynaJ), xylan beta-1,4-xylosidase (xynB), xylose rep
 ressor (xylR), xylose isomerase (xylA), xylulose kinase (xyl
 40 B), YncB (yncB), YncC (yncC), YncD (yncD) and YncE (yncE) ge
 nes, complete cds. NID: g1750106. >gp:gp|Z99113|BSUB0010_38
 Bacillus subtilis complete genome (section 10 of 21): from 1
 781201 to 2014980. NID: g2634090.
 atgcaagatttttagcaatttagttgaagaagttgaaaacacacttattccttacttttaga
 45 aaaattgaaaagcgtgcattattttaatcaggaaaaggtcttaaatgcttttcaccatggt
 aaagctagcgaaaagtgattttacaggggtctacgggttatggatatgatgattttgggaga
 gaccatttagaacaattttatgcgcacacatttaaagcagatgacgcacttgtaagacct
 caaattattttcaggtactcatgctattacttttagctttacaaagtacgttaaaaaacaat
 gatgaactacttttatattacaggtagtcctatgatatacacttctagaagtcattggtata
 50 aatggcaatggtgttgaaagtcctaaagaatatggtgttcgctataatgaagtcgaatta
 cgtgacggtcgaattgatattcctaaagtcacactgcaattaatgacaatacaaaaagt
 ttagcaattcaacgatcaaaaggatatga

Sequence 1464

MQDFSNLVEEVENTLIPIYFRKIEKRALFNQEKVLNAFHVHKASESDLQGSGTGYGYDDFGR
 55 DHLEQIYAHTFKADDALVRPQIIISGTHAITLALQSTLKNDELLEYITGSPYDTLLEVIGI
 NGNGVESLKEYGVRYNEVELRDGRIDIPKVITAINDNKSCSNSTIKRI*

Sequence 1465

Contig_0608_pos_205_948,
is similar to (with p-value 4.0e-17)
>sp:sp|P42237|GUDT_BACSU PROBABLE GLUCARATE TRANSPORTER. >gp
:gp|D30808|BACYCB20_4 Bacillus subtilis DNA around 20 degree
s region of chromosome containing yckA-T genes. NID: g709995
5 . >gp:gp|Z99105|BSUB0002_77 Bacillus subtilis complete genom
e (section 2 of 21): from 194651 to 415810. NID: g2632457.
atggcagttccttgggcaattattgctaagacttaccagaacaacataaaatggtaaac
gatgccgagaaaagatttattactcaaaatcgtgatattgtcgcaactgagaaatcttta
10 ccaccgtggaacggtttttaaagtcatttcagtttctacgcaatcgcaattgcaatacttt
gtagttcagtttgttatcgctgttctctaataatggttaccaacataatttaactgaacaa
tatcatgtgaatttcaaagaaatgactatcagtgcaattaccttggtattttatgttcttc
ttaattttatttgcctggagctatttcagacaagattttgaatacaggtcaatcacgtttt
gttgacgtggcgtaattgcgattgcgggatttgggtattctcaatttcaattttctta
15 gcagtacatacagacaacttatatgtaaccattttctggttatcactttgtttagggtggc
gtaggtattttctatgggaatgagttgggctgcagccactgacttaggtcgtaatttctct
ggaactgtgtcaggttgatgaacttatgggaaatgttggcgcaacttattagtccttta
cttgagggaattgttagatcatttaggatggtaaatgacattacaactcttaacgttt
ccggctattatagcagcaattctatggttcttcgttaaccagacaaaacctcttattggt
20 agtgaagaacacggttaataaataa

Sequence 1466
MAVLWAIIAKDLPEQHKMVNDAEKRFITQNRDIVATEKSLPPWKRFLSHFSFYAIALQYF
VVQFVIALFLIWLPTYLTEQYHVNFKEMTISALPWLFMFFLILFAGAISDKILNTGQSRF
25 VARGVIAIAGFVVFISIFLAVHTDNLYVTIFWLSLCLGGVGISMGMSWAAATDLGRNFS
GTVSGWMNLWGNVGALISPLLAGIVVDHLGWSMTLQLLIVPAIIAAILWFFVKPKPLIV
SEEHVNK*

Sequence 1467
30 Contig_0608_pos_1553_3469,
putative peptide of unknown function
atgcaacaaatcattaaactcttggatacacttcgaccgttcaaaaattgatatagctaaa
ggaatacagacaaggatttttaatgatactaccagcattgataggttacttattaggattc
cctatgtttggatttctaataatcaacaggtacgctagcacatgtctacgttttttagtgga
35 tcaccacaatctatgttaaaaacagtcacacgttgcatttactatttgcagtg
attcttggcactttaacagtatctcagcctattttatttggattactattactgattggt
gttacaatccccatattatacgttttaatgcactaaaaatcgctgggtccatcatcgacattc
tttttagtaacggttttgccttatctataaaacttaccgatagccccagaagaagcgcttta
cgtggatctgcaattctcattgggtggtatattggctaccataaacagtaattttaacaatc
40 atatttgctaaagagaaagcagaagacagagcaattcatgcggattttaaaacattacat
aacttgctacatcattttgatgagccagaggatttcaaagcatatgctcgaaacgctggt
acagaatttagaaattctgaaaaacttttaattacctcaacatcaggtggtaatggaaaa
ttaagtaaaagggttcagaaattaattttattacacacatcagcacaagggatatttca
gaattactagaactcaatgaaaatcatattcgtccattaccaagtgacttagttgaaatg
45 atggatcatatcattaatagtggttcaacaacctaaacaacaatatcgaccgtggtcaaaa
gttgttgatgtggcaccagaatttcaaaatttaattggatcatattttgaaaatagatgaa
atgattcacgcaaacgataatcaaattaaatatgaagcagatattcgcaagcctttatat
agtaagcgaatatatacaaaatctaacttgcactcaattgtattttagaaatgctttgcaa
tatacagttattatggcagtagctatatttattgctctagcgtttaacattcaaaaagcg
50 tattgggtgccattgtcagcgcataccatcatgttaggtaattgtgacaacgattcgtacg
ttagacaggtcacttgctagaggatttggaaacgataatcgggactattgtttgtcgga
atcttggcatttcatatcgatcctattttcgctattatcattatgggattttctgccatg
atgacgggaagcgtttgtggcatctaaactatgcatttgcagtcattttttatcagacacaa
gtcattatgctcaatggtctagcctctcaaaatttaaatattgagatagcgtacacccga
attattgacgttctgatagggtatagctattgcagttattgggtatattcatactcgcgct
55 aaaactgcacctcaatgttatctgatgctattgccgaattgggtacgtaaagaaggatt
ttatttcattatttttttcaaaaaataaacaggaaacgaatgaacgtgatagggtagaa
agtttaaatttaaacgtcaaaatcagtaattgtgacacagatttacaattcagcgaatggt
gaattgttttagtaataaagaagcggttaagggtattactatccaagcatatttgccttagag

gaaataagttttatgctagagcgtgccatgaataataaacatcgacaaacaataaatgat
gatttaattgggtgaatatatttagtcgtatttgaaaatatagctaagcattttccaatttcaa
gcagatttaaatctcagagacatgcaaccattacctaatacattacatccgtgcttcc
cttatgaatatacaacgtaatttgtgttgaacaacgtcaggccatcacaaaagattaa

5

Sequence 1468

MQQIINSLHFDNRKIDIAKIRQGFMLPALIGYLLGFPMFGILISTGTLAHVYVFSG
SPQSMKLTVITCSLSFTICMILGTLTVSQPILFGLLLLVVTIPYYTFNALKIAGPSSTF
FLVTFCLINLPIAPEEALLRGSAILIGGILATITVILTIIFAKEKAEDRAIHADFKTLH
10 NLLHHFDEPEDFKAYARNAVTEFRNSEKLLITSTSGGNGKLSKRFOKLILLHTSAQGIYS
ELLELNENHIRPLPSDLVEMMDHIIINSVQQPKQQYRPWSKVVDVAPEFQNLMDHILKIDE
MIHANDNQIKYEADIRKPLYSKRIYQNLTFDSIVFRNALQYTVIMAVAFIALAFNIQKA
YWPVLSAHTIMLGNVTTIRTLDRSLARGIGTIIGTIVLSGILAFHIDPIFAIIMGFSAM
MTEAFVANSYAFAVIFITQVIMLNLGLASQNLNIEIAYTRIIDVLIGIAIAVIGIFILAR
15 KTASSMLSDAIAELVRKEGILFHYLFSSKNKQETNERDRVESLNLNVKISNVTQIYNSANG
ELFSNKEAVRYYYPSIFALEEISFMLERAMNNKHROTINDLDMGEYLVVFENIAKHFOQFQ
ADLNLRLDMQPLPQYNYIRASLMNIQRNCVEQRQAITKD*

Sequence 1469

20 Contig_0608_pos_3547_3960,

putative peptide of unknown function

atgccttgagcagatggaagattatcctcaaagttggaaaaactttgaagaactcgaacgg
aaaaaagcgattgatatttggaatgcatgctcaaagatggctataaagaatctgatgtg
atacctattgctacaaatcaagctgaaaaatggatgaacatgcttcaaaagaagaatta
25 gagacgttaaaaaacaaacatatcacgcaacatcaagaagatgaatcagctaatacctaaa
cttaacgaagaaaatgttcatgtatattatgaagatcagctatggaaagtaaaatctaaa
gaggctagacgagcttcagatacatttgacacaaaatctgaagcagttaaccgtgcacaa
catatcgagagaataaaggtagcaaaagtgattgagcatcgaaaagatgagtga

30 Sequence 1470

MPWTMEDYPQSWNFEELEKKAIDIGNAMLDGKYKESDVIPIATNQAQEKWYEHASKEEL
ETLKNKHITQHQEDESANPKLNEENVHVVYEDQLWKVKSKEARRASDTFDTKSEAVNRAQ
HIAENKGTKVIEHRKDE*

35 Sequence 1471

Contig_0608_pos_12044_11130,

is similar to (with p-value 7.0e-76)

>sp:sp|P94463|FMT_BACSU METHIONYL-TRNA FORMYLTRANSFERASE (EC
2.1.2.9). >gp:gp|Y10304|BSPRIADFS_3 B.subtilis priA, def, f
40 mt, sun genes. NID: g1772497.

atgggaacacctgatttttcaacgaaaatttttagagatgttaattgctgagcatgaagtt
atcgctgtagtgcacaaacctgatagaccagtgggacgtaagaaagtgatgcaccacca
ccagtaaaaagagtagctacaaagcatcaaataccggtatatcaacctgaaaaacttaaa
gattctcaagaattagaatcggtactttcttttagaatcagatttaataagtaacggctgag
45 ttcggtcaactattaccagagtccttactcaatgcacctaataataggagctattaatgtc
catgcatcattgctacctaagtatagaggaggagcacctatacatcaggctataattgat
ggtgaagaagaaaactggaatcacgattatgtatatggttaaaaaacttgatgcaggtaat
atcatctcgcaacaatcaattcgattgaagaagaagataatgttggcgcaatgcatgat
aaattaagcttttttaggtgccgaattattaaagaagacacttcctagtatcattgataat
50 accaatgacagtatccctcaagatgatgcacttgcaacatttgcatctaataattcgctcgt
gaagacgagagagttgattggaatatgagtgcaagcaattcataaccatattagagga
ctgtctccatggccagttgcttatacaactatgaatgaaaagaatctcaaatatttttagc
gctttcattgtgaaaagggaaaaaaggtaatccaggaacaattattgaaactactaagcat
gaactcatcatagctaccggttctgatgatgccatcgacttactgagattcaacctgca
55 gggaaaaaacgtatgaaagttactgattatttaagtgggtgtacaagagtcgttagttggg
aaagttctattatga

Sequence 1472

MGTPDFSTKILEMLIAEHEVIAVVTQPDPRVGRKKVMTPPPVKRVATKHQIPVYQPEKLK

DSQELESLLSLESDLIVTAAFGQLLPESLLNAPKLGAINVHASLLPKYRGGAPIHQAIID
 GEEETGITIMYMVKKLDAGNIISQQSIRIEEDNVGAMHDKLSFLGAELLKKTLPISIIDN
 TNDISIPQDDALATFASNIRREDERVDWNMSAQAIHNNHIRGLSPWPVAYTTMNEKNLKLFS
 AFIVKGGKGNPGTIIETTKHELI IATGSDDAIALTEIQPAGKKRMKVTDYLSGVQESLVG
 5 KVLL*

Sequence 1473

Contig_0608_pos_11121_9826,

is similar to (with p-value 1.0e-92)

10 >sp:sp|P94464|SUN_BACSU SUN PROTEIN. >gp:gp|Z99112|BSUB0009_
 44 Bacillus subtilis complete genome (section 9 of 21): from
 1598421 to 1807200. NID: g2633902. >gp:gp|Y13937|BSY13937_1
 0 Bacillus subtilis genomic DNA from the spoVM region. NID:
 g2337793.

15 gtgcgaacatatgcattagaacaatcaacgacgtcctaaataaaggggcttataagtaat
 ttgaaaattaatgaagttctatctacaaataacattaatacagtagataaaaaatttattc
 acagaattagtagtatatggaacaataaaaaagaaaataactcgttagattatctactaaagcct
 tttatcaaaactaaaaatcaaatcatgggtgcgacaattactgtggatgagtttatatcaa
 tatttatatttagataaaaaatcctaaccatgctattattcatgaagcggtagatatagca
 20 aaaaaacgtgggtggctatcacacaggaatatagtcaatgggtgtattacgaacagtaatg
 cgcactgaattgccaaagctttgaagatatagatgatactaaaaaagaattgcaattcaa
 tatagtcttcccaaatggattgttgatcattgggttacacattttggagtagaaaaaact
 gaaaacattgcacgatcttttttagagcctgtaaccacaaccggtgcgcgccaatatatct
 cgtgattctattgattcaattatctctaagttagaacaggaaggttaccacgttaaaaaa
 25 gacgatattgttaccattttgtcttcatatatcaggtatgcctgtggttaattcaaacgct
 tttaaagaaggttatatctctattcaagataaaaagttcaatgatggtagcttatgtaatg
 aacctaggcgagatgacaaagtttttagatgcgtgcagcgacactgggtggttaagcttgt
 catatggcagaaattctttcaccagaaggtcacgtcgatgcaacagatattcatgaacat
 aaaataaattcttataaagcaaaatattaaaaaattgaaattgaataatatcaagctttt
 30 caacatgatgctacagaagtatacgataaaatgtatgataagattcttgttgatgcacca
 tgtagtggattaggtgttcttagacacaaaacctgaaattaaatatagtcaatcacaaaat
 agcattaaagtccttttagtagaattacaattacaatttttagaaaaatgttaagataaatatt
 aaacctgggtggtacaatagtgtattcaacatgtacaatagaacaaatggaaaacgaaaat
 gtcacttatacttttttaagagacataaagattttgagtttgaaccattccaaaatcca
 35 gcgactggtgaacaggttaaaacggttacagatacttccacaagattttaattcggatgga
 ttctttatttagcaagataaaaaagaaaggaaagtttag

Sequence 1474

VRTYALETINDVLNKGAYSNLKINEVLSTNNINTVDKNLFTLVYGTIKRKYSLDYLLKP
 40 FIKTKIKSWVRQLLWMSLYQYLYLDKIPNHAIHEAVDIAKKRGYHTGNIVNGVLRVTM
 RTELPSPFEDIDTTKKRIAIQYSLPKWIVDHWVTHFGVEKTENIARSELEPVTTTVRANIS
 RDSIDSIISKLEQEGYHVKKDDMLPFCLHISGMPVVNSNAFKEGYISIQDKSSMMVAYVM
 NLGRDDKVLDACSAPGGKACHMAEILSPEGHVDATDIHEHKINLIKQNIKKLKLNNIKAF
 QHDATEVYDKMYDKILVDAPCSGLGLVRHKPEIKYSQSQNSIKSLVELQLQILENVKONI
 45 KPGGTIVYSTCTIEQMENENVYITFLKRHKDFEFEPFQNPATGEQVKTLQILPQDFNSDG
 FFISKIKRKES*

Sequence 1475

Contig_0608_pos_8722_7979,

is similar to (with p-value 1.0e-26)

50 >gp:gp|Z70722|MLCB1770_13 Mycobacterium leprae cosmid B1770.
 NID: g2344819.
 atgttaaacgcacatttttctactgatactgggcaacatcgtgagaaaaacgaggacgct
 ggcggtatattttacaatcaaacacagcaacaaatgctagtattatgcgatggcatgggt
 55 ggacatcaagctggagaaatagctagtcagtttgtcacttatgaacttcaaaagcgtttt
 gaagaagagaatctaattgaaataaatcgtgctgaatcgtggttgcgttcgaacattaaa
 gaaatcaattttcagctgtacaactatgctcaagaaaatgaagattacagaggtatgggt
 acaacgctcgtttgtgccatcatttatgacaaacaagttgtttagcaaatgtaggagat
 tcgcgtgcttatgtaattaatcagagacagatggatcaaattacgagcgaccattcattt

gttaatcacttagtaatgactggacaaattactaaagatgaagcatttcatcatccacaa
 cgtaatatattactaaagtcattgggaacagataaacgtgttctccagatttatttattc
 aagagaactcatttttatgattatcttcttttgaaactctgacggacttactgattatgtc
 agagattatgaaatccaagaactactaagttcaaataattcattagacgtccatggtaat
 5 gagttattggacttagcgcttgcccatgattcaaaagataatgtcagctttatcctttta
 aagttagaagtgataaagtatga

Sequence 1476

MLNAQFFTDGQHREKNEDAGGI FYNQTQQQMLVLC DGMGGHQAGEIASQFVTYELQKRF
 10 EEENLIEINRAESWLRSNIKEINFQLYNYAQENEDYRGMGTTLVCAIIYDKQVVVANVGD
 SRAYVINQRQMDQITSDHSFVNHLVMTGQITKDEAFHHPQRNIITKVMGTDKRVSFDLFI
 KRTHFYDYL LNSDGLTDYVRDYEIQELLSSNNSLDVHGNE LLDLALAHDSKDNVSFILL
 KLEGDKV*

15 Sequence 1477

Contig_0608_pos_7766_5979,

is similar to (with p-value 4.0e-55)

>gp:gp|Z70722|MLCB1770_9 Mycobacterium leprae cosmid B1770.

NID: g2344819.

20 atgattgacgtagatgaggaagatgattgtttttatattgttatggaatacatagaaggt
 cctacttttagcggaatatatacacagtcattggtccacttagtgtagaaactgctattcag
 ttacagaacaaattttgagtggaatcaaactgcgcattgatagagaattgttcatcgc
 gatattaaaccacagaatatattaattgataagaataaaaaattacaaatttttgatttt
 ggtattgctaaagcattgagtgaaacgtcgctgactcaaacaaatcatgttttaggaact
 25 gttcaatatattatcgctgaacaggctaaaggcgaagctactgatgaaagtactgacata
 tattcaattggaattgtattatagagattgttagtgagagccaccattcaatggagag
 actgcagtgagttattgctataaaagcatattcaagatagttatcccaaatataacaacggat
 aaaagagatgatgtacctcaatcattgagtaattgtcgttttacgtgctaccgagaaagat
 aaatatcatcgctatcatactgttcaagaaatgtgtgacgatttaacaagtgcctttacat
 30 gagaatcggttgatgaagaaaaatcacgagctagataaaactaaaacgggtaccactcact
 aaagatgatttgaatcataaaatctatgatgaacaaaatcaaaatgaccttaataaaacc
 atgcaaatacctattgttaacgattcaataaaagcaacaagaatttcaatcgctctgaacca
 cggttattatcaaaaagcgacaagaacggttctaaaatgaaaattgcaattttatcaatc
 atttttgtaattattattattgtttttcttttgtagctatggctgtttttggaat
 35 aaatatgaagaaatgcctgaccttaaagggaactgaaaaacaagctgaaaaggattta
 gacagtcattcatctaaaagtaggtgacatatcaagaaattatagtgataaatatcctgaa
 aacaaattattaaaacaaatccagatagtgagaaacgcgtcgaaacagggaatagaatt
 gatatcggtctatccaaaggaccagagaaggttaagatgccaaatgttttaggtatgtcg
 aaaaatgatgcgctaaaaaaattaaaggctatcggtttaaagatattcacgttgagcaa
 40 gcttatagtaacaacatgaaaaaggattatttctgaacaaagcgttgtagctaattagt
 gaggttgccgttaataatcatcatattacaatttatgaatcattaggtgttcgacaagta
 tatgtcaataaattatgaaaataagtcattatgagtcagcaaaaaaagaacttgaagataaa
 ggatttaaaagtcaagtgacaaaagaaaacacgatgatgtcgaaaaaggtaattgtcatt
 tctcaatctccaaaggalaaaactgttgatgaaggttctactatactattagtggtttct
 45 aaaggagaaaagtctgaagaagaagatgatgaggaggacaaggatacaacgactaaaaat
 gagactgttaaagtaccgtataccggtaaaaaaagtaaaagtcaaaaagtagaagtattt
 attcgtgatattgaaaataaagggtgattcagcagttcaaacgtttaatattaaaagtgat
 aaaacaattaatattcctttgaaaattaaaaaagggaagtgcgctgggtacaccataaga
 gttgataataaaattgtagctgataaagatgtgagctatgatggctaa

50

Sequence 1478

MIDVDEEDDCFYIVMEYIEGPTLA EYI HSHGPLSVETAIQFTEQILSGIKHAHDMRIVHR
 DIKPQNILIDKNKKLQIFDFGI AKALSETSLTQTNHVLGTVQYLSPEQAKGEATDESTDI
 YSIGIVLYEMLVGEPPFNGETA VSI AIKH IQDSIPNITTDKRDDVPQSLSNVLRATEKD
 55 KYHRYHTVQEMCDDLTSALHENRLNEEKYELDKTKTVPLTKDDLNHKIYDEQNQNDLNKT
 MQPIPVNDSIKQEQFSSEPRYYQSKDKR SKMKIAILSII FVILLIGLFSFVAMAVFGN
 KYEEMPDLKGKTEKQAEKVLDSHHLKVGDISRNYSDKYPENQIKTNPD SGERVEQGNRV
 DIVLSKGPEKVKMPNVLGMSKNDALKK LKAI GFKDIHVEQAYSQTYEKGLISEQSVVANS
 EVAVNNHHITIYESLGVRQVYVNNYENKSYESAKKELEDKGFVKVQVTKENND DVEKGNVI

SQSPKDKTVDEGSTILLVVSKEKSEEEDEEDKDTTCKNETVKVPYTGKKSKSQKVEVF
 IRDIENKGD SAVQTFNIKSDKTINIPKIKKGS DAGYTIRVDNKIVADKDVSYDG*

Sequence 1479

5 Contig_0608_pos_5685_4795,
 is similar to (with p-value 4.0e-39)
 >sp:sp|P45339|YJEQ_HAEIN HYPOTHETICAL PROTEIN HI1714. >pir:p
 ir|B64176|B64176 hypothetical protein HI1714 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32844|U32844_6 Haemophilu
 10 s influenzae Rd section 159 of 163 of the complete genome. N
 ID: g1574563.
 atgagaggtgtcttttgaagactgggtcgaatcggttaactaatcagtggtgtgtatcaa
 gtacacgtagaaggtagagatttgataccaaaccacgtggtttattcagaaaaaagaag
 ttttcacctgtggtggcgatcgtagattttgaagttcaaaatacaaaagagggctat
 15 attcatcatgtacatgaccgaaataatgaactaaaacgaccacctgtaagtaattattgac
 gagttagttatagtaaatgagtgagtcgagcctgaattttcaacacaattattagatcgc
 tatttagtgattgctcattcttatcatctcaaacctagaatttttaactactaaacatgat
 ttagcttccgaacaagaaattcttaaaatcaaagacacaataaaaatatatcaaaaaata
 ggtatgctacgcagtttattggaaaagatagtaattatactgctactgttgatgaatgg
 20 tctgacggtttaatagtattaaagtgccaatctggagtggttaaatctactttcttaaat
 agttatcagcctcagttgaagttagaacaaatgatatttctaagtcattgaataggggt
 aaacatactacaagacatgtcgaattatácgatágaaggggtgttacatcgctgataca
 ccgggggttagtgcggttagatttttaatacatattgaaaaagaacaactaaaagatttttt
 attgatattcatgaagctggagagcaatgtaagtttcgtaattgtaatacaaaaagaa
 25 ccacaatgtcatgtcaaagcactcggtgaaaaaggagaaattccacaattcaggtatgat
 cattatcagcaattatataatgaaattttccaatagaaggttcgatactaa

Sequence 1480

MRGVFLKTGRIVKLISGVYQVDVEGERFDTKPRGLFRKKKFSPVVGDIVDFEVQNTKEGY
 30 IHHVHDRNNELKRPPVSNIDELVIVMSAVEPEFSTQLLDRLYLVIHSHYHLKPRILITKHD
 LASEQEILKIKDTIKIYQKIGYATQFIGKDSNYTATVDEWSDGLIVLSGQSGVGKSTFLN
 SYQPQLKETNDISKSLNRGKHTRHVELYDRKGGYIADTPGFSALDFNHIEKEQLKDFP
 IDIHEAGEQCKFRNCNHIKEPQCHVKALVEKGEIPQFRYDHYQQLYNEISNRKVRY*

35 Sequence 1481

Contig_0609_pos_4636_3806,
 is similar to (with p-value 2.0e-88)
 >sp:sp|P18156|GLPF_BACSU GLYCEROL UPTAKE FACILITATOR PROTEIN
 . >pir:pir|C47700|C47700 glycerol uptake facilitator glpF pr
 40 otein - Bacillus subtilis >gp:gp|M99611|BACGLPPEK_2 Bacillus
 subtilis antiterminator regulatory protein (glpP), glycerol
 uptake facilitator (glpF) genes, complete cds, glycerol kin
 ase (glpK) gene, 5' end. NID: g142995. >gp:gp|Z99108|BSUB000
 5_196 Bacillus subtilis complete genome (section 5 of 21): f
 45 rom 802821 to 1011250. NID: g2633055. >gp:gp|Z99109|BSUB0006
 _3 Bacillus subtilis complete genome (section 6 of 21): from
 999501 to 1209940. NID: g2633260. >gp:gp|Y14079|BSY14079_3
 Bacillus subtilis chromosomal DNA, region 75 degrees: glpP
 D operon and downstream. NID: g2226133.
 50 atgtatatgaatgcttatttagcagaatttttaggtactgcaatccttattcttttgg
 ggtggcggtttgtgcaaacgttaacttaagagaagtgctggttaacggtgcagattggatt
 gttattgcatttggttgggtttggcagtaacaaatggcggtttatgctgttggaacgttt
 tctggtgcacatttaaatccagctgtaacagttgctttagccatggatggtggatttagc
 55 tggcgcaagtaccgggctatattgtttgtcaaatgcttgccggtattgttggtggagt
 tttgtatgggttaattgtatttaccacactggaaagttacagaagatccagcagtc aaatta
 ggtgtattttcaacagcaccagccattaaaaattattttgctaactttttaagtgaagatt
 atcgggactatggctttaacattaggaattttatttatcgggggttaataaaattgctgat
 ggttttaaatccaattattgttggttagtcttatcatagcaattgggttaagcttaggaggt
 actactgggttacgctattaatccagcccgtgacctagcaccacgtattgcacatgctatt

ttgccaatcatggttaaaggtaaatctaactggtcttacgcaattgtacccgttctggga
cccatggcaggtggtatgttaggtgcgattgtttacgaagtgtttataaacaacattc
aatttttagttgtttcattggtttaattgtacttatattcacacttatacttggcgtgata
ctaaataagatatctcaaaaataaaaacaacgatattgaatcaatttattaa

5

Sequence 1482

MYMNAYLAFLGTAILILFGGGVCANVNLKRSAGNGADWIVIAFGWGLAVTMGVYAVGTF
SGAHLNPAVTVVALAMDGGFSWAQVPGYIVCQMLGGIVGGVFVWLMYLPWKVTEPAVKL
GVFSTAPAIKNYFANFLSEIIGTMALTLGILFIGVNKIADGLNPIIVGSLIIAIGLSLGG
10 TTGYAINPARDLAPRIAHAILPIHGKGSNWSYAIVPVLGPMAGGMLGAIVYEVFYKQTF
NFSCFIGLIVLIFTLILGVILNKISQNNNDIESIY*

Sequence 1483

Contig_0609_pos_3652_2153,
15 is similar to (with p-value 0.0e+00)
>sp:sp|P18157|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:
GLYCEROL 3-PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK). >pir:pi
r|B45868|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus sub
20 tilis >gp:gp|M34393|BACGLPKD_2 B.subtilis glycerol kinase (g
lpK) and glycerol-3-phosphate dehydrogenase (glpD) genes, co
mplete cds. NID: g142990. >gp:gp|Z99108|BSUB0005_197 Bacillu
s subtilis complete genome (section 5 of 21): from 802821 to
1011250. NID: g2633055. >gp:gp|Z99109|BSUB0006_4 Bacillus s
ubtilis complete genome (section 6 of 21): from 999501 to 12
25 09940. NID: g2633260. >gp:gp|Y14079|BSY14079_4 Bacillus subt
ilis chromosomal DNA, region 75 degrees: glpPFDK operon and
downstream. NID: g2226133.
atggaaaaatataatatttatcaattgatcaaggaaactacgagttcacgtgcgatactttt
aataaagaaggagaaattaaagggtgtttctcaagagaatttaacaacactttccacat
30 ccaggctgggtagaacatgatgctaataatgaaatattggacatctgttctatcagttatggct
gagttacttaataaaaacaatattaatgcaaatcaaattgaaggtattggtattacaaac
caacgtgaaacgacagttgtatgggataaaaaatacaggtcgtccaatctatcacgctatc
gtttggcaatcacgtcagacacacaagatatattgtacaaatttaaggaacagggttatgaa
gaaacatttagagaaaaaacagggtttacttttagaccgctactttgcggaactaaagta
35 aaatggattcttgatcatgttgaaggtgctagagaaaaagctgaaatggtgatttactc
ttcggaaacatcgattcatggttagtatggaaattgtcaggacgtactgctcatattaca
gattacactaatgcaagtcgtacattaatgtttaatatttatgacctaaaatgggatgat
gagttgttagaacttttaaatattcctaacaacaaatgttacctgaagttaaagaatcaagt
gaaatttaccgggaaaactatcgactatcacttctttggtcaagaagtacatttgcgtggt
40 attgccggtgaccaaacagcagcattatttgggtcaagcatgtttgaccgtggtgatgta
aaaaatacatcacggcacaggtggatttatgctaataactggtgaagaagcagttatg
tcagaaagtggcttggtaacaaccattgcatacgggttagatggaaaagttatattgca
cttgaaggttcaattttcgtatctggttctgctatccaatggctacgagatggtttgaga
atgattaattctgcccacaaaccgaaaactatgcttcaagagtagagtcactgaggggt
45 gtttatatggttccagcatttgttgggttaggtacaccttattgggattcagaagcaaga
ggtgctattttcggattatctcgtggtacggaaaaagaacatttcattcgtgctacatta
gaatctttgtgctatcaaaacaagagatgttatggaagctatgtctaaggactcaggtatt
gaagttcaaaatttacgcgttgatggtggtgctgtaaaaaataacttcattatgcagttc
caagcagatatcgtaaatctctgttgaaagacctgaaatccaagaaacaacagcactt
50 ggtgctgcatatttagctggattagctgttggtgattctgggatgataaagaggatccgt
gaacgttggaacttcaaaactgagttcaaaaccagaaatggatgcagatcaacgtcataaa
ctttatagtggttggaaaaaagctgttaaggcgactcaagtatttaaatagaagattaa

55 Sequence 1484

MEKYILSIDQGTSSRAILFNKEGEIKGVSQREFKQHFPHPGWVEHDANEIWTSVLSVMA
ELLNENNINANQIEGIGITNQRETTVVWDKNTGRPIYHAIWQSRQTQDICTNLKEQGYE
ETFREKTGLLLDPYFAGTKVKWILDHVEGAREKAENGDLFGTIDSWLVWKLSGRTAHIT
DYTNASRTLMEFIYDLKWDELLELLNIPKQMLPEVKESSEIYGKTIIDYHFGQEVPIAG

IAGDQQAALFGQACFDRGDVKNITYGTGGFMLMNTGEEAVKSESGLLTTIAYGLDGKVN
LEGSIFVSGSAIQWLRDGLRMINSAPOQENYASRVESTEGVYMPAFVGLGTPYWDSEAR
GAIFGLSRGTEKEHFIRATLESICYQTRDVMAMSKDSGIEVQNLVDDGGAVKNNFIMQF
QADIVNSSVERPEIQETTALGAAYLAGLAVGFWDKEDIRERWKLQTEFKPEMDADQRHK
5 LYSGWKKAVKATQVFKLED*

Sequence 1485

Contig_0609_pos_1976_303,

is similar to (with p-value 0.0e+00)

10 >sp:sp|P18158|GLPD_BACSU AEROBIC GLYCEROL-3-PHOSPHATE DEHYDR
OGENASE (EC 1.1.99.5). >pir:pir|C45868|C45868 glycerol-3-pho
sphate dehydrogenase (EC 1.1.99.5) - Bacillus subtilis >gp:g
p|M34393|BACGLPKD_3 B.subtilis glycerol kinase (glpK) and gl
ycerol-3-phosphate dehydrogenase (glpD) genes, complete cds.
15 NID: g142990. >gp:gp|Z99108|BSUB0005_198 Bacillus subtilis
complete genome (section 5 of 21): from 802821 to 1011250. N
ID: g2633055. >gp:gp|Z99109|BSUB0006_5 Bacillus subtilis com
plete genome (section 6 of 21): from 999501 to 1209940. NID:
g2633260. >gp:gp|Y14079|BSY14079_5 Bacillus subtilis chromo
20 somal DNA, region 75 degrees: glpPKD operon and downstream.
NID: g2226133.

atgtcattatctacattgaaaagggatcatattaaaaagaatttaagagacactgaatac
gatgttggtatcgtaggtggcgttattacaggtgcaggtattgcttagatgcaagtaaat
cgtgggatgaaggtagctttagtagagatgcaagactttgcacaaggtacaagttcacgc
25 tcaactaaactgtacacggtggtttagatatttaaaacaactgcaagtagggtagtt
gcagaaacaggtaaagaacgtgctattgtttatgaaaatgggtccacatgtgacaacacca
gaatggatgcttttacctatgcataaaggtggtacatttggtaaattctcaacttctatt
ggactagctatgtacgatcgcttagctggtgtcaaaaaatccgaacgtaaaaaaatgta
tctaagcaagaacggttaaataaagaacctttagttaaacgtgatggattaaaagcggt
30 ggctactatgtggaataccgcactgatgatgcgcgtttaactattgaagttatgaaaaa
gctgctgaaaatggagcagaaatcattaattatacaaaatcagaacacttcacttatgat
tccaataaagaagtaaatggtattgaagatttgatattgattgagcgaaacgtatgag
attaaagctaaaaaagttattaatgcttctggtccttgggttgatgaagtgaagtggtg
gattatgcacgtaacaataagcaattaagattaactaaaggtgtacacggttggtatagat
35 caatctaaattcccattaggtcaagcagtttactttgatactgaaaagacggacgcatg
atttttgcgattccacgtgaaggaaaagccttatgttaggaacaactgacacgttttatgat
aatgaaaaagcaacacctttaacaacacaagaagatagagactacttaattaatgcaatt
aactatatgttcccaacagttaatgttaagatgaagatattgaatcaacatgggctggt
attcgctccgctaattcttgaaaaggtaaaagatccttctgaaatctcacgtaaaagatgaa
40 gtttggaaggtgaatctggattattaactatagcaggcggtaaattaactggttatcgt
catatggcactagaaattgttgatttattagctaaacggttaaaacaagaatacggattg
aaatttgatcatgtgccacaaaaaatctaaaaatttccggtggtgacggtggcggaagc
aaaaactttgaacactttgtgaacaaaaagttgatgcagctaaaggatttggaattgat
gaagatgtggcagctcgcttagcaagtaaatatggttcaaattgttgatcaactattta
45 attgctcaaacggcaccatatcatgatagtaaataccattagaaatttatgttgaaatta
gtttatagatttcaacaagaatgggtttacaaaccaactgacttcttagtacgtcgttct
ggcaaatatactttaatttcaagatgtgttagattataaaaatgctgtgatagatgtt
atggcggtatgcttaattatagtgaactcaaaaagaagcttatactgaagaagtagaa
gttgcgattgatgaggcacgtacaggtaatgatcaacctgcaactaaagcttaa

50

Sequence 1486

MSLSTLKRDIKKNLRDTEYDVVIVGGITGAGIALDASNRMKVALVEMQDFAQGTSSR
STKLHVHGLRLYLKQLQVGVAETGKERAIYENGPHVTTPEWMLLPMLIKGGTFGKFSTSI
GLAMYDRLAGVKKSERKKMLSKQETLNKEPLVKRDGLKGGYYVEYRTDDARLTIEVMKK
55 AAENGAIEIINYTKSEHFTYDSNKKVNGIEVLDMIDGETYAIKAKKVINASGPWVDEVRSG
DYARNNKQLRLTKGVHVIDQSKFPLGQAVYFTEKDGRMIFAIIPREGKAYVGTDTDFYD
NEKATPLTTQEDRDYLINAINYMFPTVNVKDEDIESTWAGIRPLILEKKGDPSEISRKDE
VWEGESGLTTIAGGKLTGYRHMALIEVDLLAKRLKQYGLKFESCATKNLKSISGDDVGGG
KNFEHFVEQKVDAAKGFIDEDVARRLASKYGSNVDQLFNIAQTAPYHDSKLPLEIYVEL

VYSIQQEMVYKPTDFLVRRSGKLYFNIQDVLVDYKNAVIDVMADMLNYSETQKEAYTEEVE
VAIDEARTGNDQPATKA*

Sequence 1487

5 Contig_0610_pos_767_1225,
putative peptide of unknown function
atgacagactcaaagtctaaagaataagaactggacgtttaattgcgataagttcatta
gtgttttgtattttacttatcatcacaccactttattgtattagatgaatcaacagctaaa
tcaattttatcttttagctggtcaaaaaacatcagatacagcagtgaaaaacatttttaaat
10 agtgaccgatacactggaattatgtatatttttagcttacttagcaggtactgttgctttc
tggaatcgccatccatattttatggtggtttatgtttgccgtatatatttctaagtcacta
tttacactcgtaaatctttacttatttattcaaggtatttttagatgtaaaaaatgtactt
gcagttttaccaattttaattgtagtgttggtctataattctagcaatttatatgcta
gttgttttctattacacgtaaaagtactttcaatagatag

Sequence 1488

MTDSNAKEIRTGLRIAISLVFCILLIIHHFIVLDESTAKSILSLAGQKTSDTAVKNILN
SDRYTGIMYILAYLAGTVAFWNRHPYLWWFMFAVYISNALFTLVNLYLFIQGILDVKNVL
AVLPILIVVIGSIILAIYMLVVSITRKSTFNR*

Sequence 1489

Contig_0610_pos_2605_3345,
is similar to (with p-value 4.0e-84)
>sp:sp|Q06174|EST_BACST CARBOXYLESTERASE PRECURSOR (EC 3.1.1
25 .1). >pir:pir|JC1374|JC1374 carboxylesterase (EC 3.1.1.1) -
Bacillus stearothermophilus (strain IFO 12550) >gp:gp|D12681
|BACPBH7_1 Bacillus stearothermophilus esterase gene. NID: g
216313.
atgcaaattaaactacaaaaccattcttttttgaagaagggaaacgtgcagtggtactt
30 cttcacggctttacaggttaactctgctgatgtaagacaacttggcggttatcttcaaaaa
aagggtatatacatcttatgctccacaatatgaaggacatgcagcgccccagaagaaata
ttaaactctagcccttttgttgggtttaaagatgttttagatggttatgattatttagta
gatcaaggttacgaagaaatagcagtagctggtttatcattaggtggcgcttcgcatta
aaactaagtttaaatcgtgatgtgaagggattataactatgtgtgcacctatggagaat
35 aaacagaaggttcgatttatgaaggctttctgaatatgcacgtaactttaaaaaatat
gaaggcaaagatcaacaaacgattgatcaagaaatggaacaatttcatccaactgaaacc
ctgagagaactgagtgacactctaaatggagtttaagaacatgtcgatgaagtaattgat
ccaatacttgcgtacaagcagaacaagatacaatgattgatcctcaatcagcaaatat
atatataatcatgtcgtattctgatgaaaaagaaatcaaatggtatcaacattcaggtcat
40 gtgattaccattgataaagaaaaagagaaagtctttgaagatgtatatcaatttttagaa
tcattggaatggacagagtaa

Sequence 1490

45 MQIKLPKPFEEEGKRAVLLHGFTHGNSADVRQLGRYLQKKGYTSYAPQYEGHAAPPEEI
LKSSPFVWFKDVLVDGYDYLDVQGYEEIYAVAGLSLGGAFALKLSLNRDVKGIIITMCAPMEN
KTEGSIYEGFLEYARNFKKYEGKDQQTIDQEMEQLFPTETLRELSDTLNGVKEHVDEVID
PILVVQAEQDTMIDPQSANYIYNHVDSDKEIKWYQHSQHVITIDKEKEKVFEDVYQFLE
SLEWTE*

Sequence 1491

50 Contig_0610_pos_3380_5758,
is similar to (with p-value 0.0e+00)
>sp:sp|P44907|VACB_HAEIN VACB PROTEIN HOMOLOG. >pir:pir|G640
98|G64098 virulence associated protein homolog (vacB) homolo
55 g - Haemophilus influenzae (strain Rd KW20) >gp:gp|U32767|U3
2767_9 Haemophilus influenzae Rd section 82 of 163 of the co
mplete genome. NID: gl573868.
atgaatttaaagcaatccatcgaagaaatgataaaacaacctgactatgaacccatgtca
gtatctgactttcaagatgcggttaggtttaaacagtgccgactcatttagagatttaatt

aaaatactcgttgaattagaacagtctggtttaattgaacgtacaagaacagacagatat
 caacgtaaacatccaacaaaacaaattcaaaactaatcaaaggaacgttaagtcaaaat
 aaaaaaggctttgctttcttaagacctgaagatgacgagatggatgatattttattcca
 ccaactaaaatcaatagagcatttagatggagatactgtcatcgtggaaattcaaaaatct
 5 cgtggagaacataaaggtaaaattgaaggtgaagtaaaatctattgaaaagcattcagtt
 acacaagttgttgaacgtatagcgaagcaaacgattttggtttcgtattaccggatgac
 aaacgtattatgcaagatatctttatacctaaaggacaaaatttaggtgctgtagatggt
 cataaagtattagtacaaattacgaagtatgccgatagtagtacaatccagaaggccac
 gtctcagcaatattaggtcataaaaatgatccaggtgtagatatactttccatcatttac
 10 cagcatggaatagaatcgagtttccagatgatgtattacaagaagctgaagaagtaccg
 gatgtaatagaaccatctgaaatcgaagggcgtcgtgatttaagagatgaattaacaatc
 actatagatggcgagatgctaaagatttagatgatgccattgctgtaaaaaaattaaaa
 aatggcaacaccgagcttacagtttagtattgcagatgtaagttactatgtaaaagaagga
 tcagcttttagataaagaagcttatgatcgtgcgacaagtgtgtatcttgcgatcgagta
 15 atcccgatgattccacaccgtctaagtaaatggaatatgctcattaaatccagaagaagat
 cgtttaacattaaagttgtcgaatggaaataaatgaacgaggcgaagttgtaaaacatgaa
 atctttgatagtgtaatatcattcaaaactacagaatgacatatgatgcagttaacaaaatt
 atcactgatcaagattctgaaatcagttcacatataaagatttaacacctatgttagat
 tttagcgaagatttatcaaatagatttaattcatatgcgcaaacgtcgtggagaaattgat
 20 tttgatattaatgaagcgaaggtacttgtgaatgacgaaggtattccaacagaagtgtta
 atgagagaacgtggcgaaggagaacgtttaattgaatcattcatgttagcagccaatgaa
 acagtagctgaacacttcaataaattggaagtaccattttatctcgtgttcataacaa
 ccaaaatctgaccgattaagacagttcttcgactttattaccaatttcgggtattatgata
 aaaggtacaggtgaagatattcatccaacaacattacaaagcattcaagaagaagttgaa
 25 ggtagaccagaacaaatgggtatttcaacgatgatgttacgttctatgcaacaagcacat
 tatgatgatgttaatttaggacattttggtttgtctgctgagtactatactcactttacg
 tctccaatagccggttatcctgatttaacagtgcatagattaattcgtaaatatttaata
 gagaattctatggataaaaaagaaatcagtcattgggaagagacgttgccagaattagct
 gagcacacatcacacgtgaacgccgtgccattgaagccgaacgtgatactgatgaattg
 30 aaaaaagctgagtatatgattcaacatattggtgatgaatttgaaggtatcatttagctcg
 gttgctaattttggtatgtttatagaattacctaataactattgagggatggttcatac
 gtaatatgacagacgattattatcattttgatgaacgacaaatggcactaatcgggtgaa
 cgtcaagcgaaggtctttcgcatgtgtgatacggcgaagttaaagtgaacacatgttgat
 gtggaagaacgcatgatagatttccaaattgttgcatgccattacctaaaaagacatca
 35 tcacaacgacctgctcgtgagaaaaccattcaagctaaaacacgtggcaagtcgttagac
 cactactaaaagtgatcgtaattggtaaaggtaaaaagaaaaaacgttaagcaacgtaaaggt
 aaaaatgcacgtaaaaaagataaacaaggtaatacgcatacaaaacctttttataaagat
 aaaagtgtagaagaagaatcgcgctgaagaaaaaataag

40 Sequence 1492

MNLKQSIEMIKQPDYEPMSVSDFQDALGLNSADSFRDLIKILVELEQSGLIERTRTDRY
 QRKQSNKTNKLIKGTLSQNKKGFAFLRPEDDEMDDIFIPPTKINRALDGDTVIVEIQKS
 RGEHKGKIEGEVKSIEKHSVTQVVGTYSEAKHFGVLPDDKRIMQDIFIPKGQNLGAVDG
 HKVLVQITKYADSTDNPEGHVSAILGHKNPDGVDILSIYQHIGIEIEFPDDVLQEAEVVP
 45 DVIEPSEIEGRRDLRDELITIDGADAKLDDAIAVKKLKNNGNTELTVSIADVSYVKEG
 SALDKEAYDRATSVYLVDRVIMIPHRLSNGICSLNPEEDRLTLSCRMEINERGEVVKHE
 IFDSVIHSNYRMTYDAVNKIITDQDSEIRSQYKDLTPMLDLAQDLNRLIHRKRREGEID
 FDINEAKVLVNDEGIPTVLMRERGERGERLIESFMLAANETVAEHFNKLEVPFIYRVHEQ
 PKSDRLRQFFDFITNFGIMIKGTGEDIHPTTLQSIQEEVEGRPEQMVI STMMLRSMQQA
 50 YDDVNLGHFGLSAEYYTHFTSPIRRYPDLTVHRLIRKYLIENSMDKKEIRHWEETLPELA
 EHTSQRRRAIEAERDDELKKAEMYIQHIGDEFEGIISSVANFGMFIELPNTIEGMVHI
 ANMTDDYYHFDERQMALIGERQAKVFRIGDTVKKVTVHVDVEERMIDFQIVGMPLPKKTS
 SQRPAREKTIQAKTRGKSLDHTKSDRNGKGGKKKKRQKRGKNARKKDKQGNTHHKPFYKD
 KSVKKKSRRKKK*

55

Sequence 1493

Contig_0610_pos_5783_6253,

is similar to (with p-value 2.0e-48)

>sp:sp|P43659|SMPB_ENTFA SMALL PROTEIN B HOMOLOG. >gp:gp|M90

060|STRATPASEA_1 Streptococcus faecalis H+ ATPase a (atpB),b (atpF),c (atpE),alpha (atpA), beta (atpD),gamma (atpG),delta (atpH),and epsilon (atpC) subunits, complete cds. NID: gl5 3565.

5 gtggctaagaaaaaatcaaaatcaccagggtacgttagctgaaaatcgtaaagcaagacat
gactacaatattgaagatacaattgaagcgggtattgctttaagagggtactgaaattaaa
tctatacgtcgtggttagtgccaatttaaaagatagctttgcgcaagtgaagcaggcgaa
atgtacctgaataatatgcatattgcaccatatgaagaagggaaccggttttaaatcatgac
cctttacgtacacgtaaaattactcttgacaaaaaagaaattcaaaaattaggtgagcgt
10 acacgagaaatagggtattctattattccggtgaagttatatttaaacatggtcaatgt
aaagttttattaggcgttgctagaggtaaaaaagaaatacgacaaacgtcaagcacttaa
gaaaaagcggtaaaacgagatattgatcgcgagttaaagcccgttattaa

Sequence 1494

15 VAKKSKSPGTLAENRKARHDYNIEDTIEAGIALRGTEIKSIRRGSANLKDSFAQVRRGE
MYLNNMHIAPIYEEGNRFNHDPLRTRKLLHKKKEIQKLGERTREIGYSIIPKLYLKHGQC
KVLGVARGGKKYDKRQALKEKAVKRDIDRAVKARY*

Sequence 1495

20 Contig_0612_pos_2290_2928,
is similar to (with p-value 2.0e-20)
>gp:gp|AJ007319|LMO34616_5 Listeria monocytogenes ascB, inlG
, inlH, inlE, dapE genes. NID: g3980132.
atgccacacctaggtacaaatgctggtgatatttttagttgattttgtaaataaatgaaa
25 caagaatataaaaaatattaaagaacatgataaagtacacgagttagacgctgttccaatg
attgagaaacatctccacagaaaaattggtgaagaagaatcacatatctactctggattt
gtaatgttaaactctgtattcaatggtggtgaacaagttaattctgttcctcataaagcg
acagctaaatataatgtaagaactgttccagaatatgacagtagctttcgtgaaggattta
tttgaaaaagtcattcgtcatgtggcggaagattatttaactgtagatatacctagcagt
30 cagcatccagtggaagtgatcggtgataatcctctattcaaaatattacacgtattgca
ccgaattatgtacatgaagacattgttggtgagtgcatgattggtacaactgatgcattc
agtttcttaggaacaaatgaaaataacgtggattttgctgtctttggacctggtgaatct
attatggcgcatcaagttgatgaatttattagaaaagatgtattttaagttacatcgat
gtttataaagatgtattttaaagcatatctagaaaaataa

35

Sequence 1496

MPHLGTNAVDILVDFVNEMKQEYKNIKEHDKVHELDAVPMIEKHLHRKIGEEESHIYSGF
VMLNSVFNGGKQVNSVPHKATAKYNVRTVPEYDSTFVKDLFEKIRHVGEDYLTVDIPSS
HDPVASSDRNPLIQNITRIAPNYVHEDIVVSALIGTTDASSFLGTNNNVDFAVFGPGES
40 IMAHQVDEFIRKDMYLSYIDVYKDVFKAYLEK*

Sequence 1497

Contig_0612_pos_9229_10425,
putative peptide of unknown function
45 atggttaaaatttatacactgtgctgatttgcatattggacagtcctttcaaatctaaaagt
tatcttagtccaaatatttttgaagatgtccaaaagagtgcatatgaaagttttaaaac
atagtcgacttagcttttaaacaggaagtcgattttattattatagcagggtgatttattt
gatagtgagaatcgtagattgctgctgaagtctttttaaataaacaatttgaaagatta
agaaaagaacaaatatttgtttatatttgccatggcaaccacgatcctcttacttctaaa
50 ataacaagtcagtgccataaacgtatccgtattttcaaatcaagtagagacatatcaa
gctatcactaaatcaggagaaacaatttatattcatggattcagctatcaaaatgatgag
agttatgaaaataaaatagacgcatacccatcaagtcaggtcagaagggcatacatatt
ggtgtattacatggaacttatagtaaatcttcggtgaaagaccgttatactgaatttagg
ttagaagacttaaatcaacgttttataccactactgggcattaggacatatcaccaacgt
55 gaacagttaagtgcacatgccagtcattaactatccaggtaatatccaaggaagacatttc
aatgaattaggagaaaaaggttgctctatttggtcgaaggtgatcatcttaactcactaca
caattttatcctactcaatttattaaatttgaaagctacaattgaaactgatcataca
tctaaacaaggactttatgatgttattcaatcttttaagataaagtaagaactgaaggg
aaatcattttatagattgaacgtacgcattaatagtgaaagacattattgcaccacaagat

ttaattcaattaaaagaatgattactgagttcgaagaaaacgaaatcaatttggtttt
 attgaagatttaaatcttcaatatgttcaaatgacgaaatgccaatagttaagagttt
 tcaccagaattacttgatgatgcgtcactttttgattcggcaatgactgatttatctt
 aatccaagggtcttctaagtttttagatgactataatgaatttgataaagttgagttagtc
 5 aatcatgcagaaagacttttaaggtgaaatgagaggtgaacaaaatgataattaa

Sequence 1498

MVKFIHCADLHLDSPFKSKSYLSPNIFEDVQKSAYESFKNIVDLALKQEVDFIIIAGDLF
 DSENRTLRAEVFLNEQFERLRKEQIFVYICHGNHDPLTSKITSQWPNNVSVFSNQVETYQ
 10 AITKSGETIYIHGFSYQNDASYENKIDAYPSSQGQKGIHIGVLHGTYSKSSVKDRYTEFR
 LEDLNQRLYHYWALGHIHQREQLSDMPVINYPGNIQGRHFNELGEKGCLLVEGDHLKLT
 QFYPTQFIKFEEATIEQDHTSKQGLYDVIQSFKDKVRTEGKSFYRLNVRINSEDIAPQD
 LIQLKEMITEFEENENQFVFIEDLNLQYVQNDempiVKEFSPELLDDASLFDSAMTDLYL
 NPRASKFLDDYNEFDKVELVNHAERLLKDEMRGEQNDN*

Sequence 1499

Contig_0612_pos_11117_13354,
 is similar to (with p-value 0.0e+00)
 >gp:gp|U21636|SAU21636_1 Staphylococcus aureus cmp-binding-f
 20 actor 1 (cbf1) and ORF X genes, complete cds. NID: g710420.
 atgcatgagcaaaaacaaaagaggttgctctacacgatcaaacacaagaatggaaaagg
 ttagaacagtcgcttaatatagagcctataaattttcctgaaaagggatagatagatac
 gaaactgctaaatctcacaacaatcacttgaacgagataaaaagtttgcgagaagaaaga
 ttaagcatattaaataaagaggcgagtcctcaatccagtagaccaaaaagtatatattgat
 25 tcgtttaatagcctttatcaacaagagactgaaattaaacaaaaagaatttgagttacgt
 tcaattgagagaagaatattgctgataagcaacgtgaactagaagctcttcaatctaata
 ggttggaagaagtggtttacgacacagacagactgaagcgatgaaaagtcataatgagt
 gatttagtattaggcaagcaagaacaaattgcttatcaatcagttagaacgtggactt
 gaagaaaataaaattgaaagaaactctaattctaattgagattaatcaagttgagaatgag
 30 cttgttcctgacgaaacctttgaaaagaaaaaggaatatacacaacaagttttagaatta
 catgaaaagagaacttgatgaaaagttaaaagaaacttttgaaagaagaacaaacacaa
 aaaaaataaaagacaaaagttttgagaataggatttattgttttgactattctatcagca
 gcactttctatttttttcttttctcactgcaaatcttatttttggtataatatttgctcta
 ttaactgtgattttttgtagtaggtatcattttttctagatctaaagcagtagattatagc
 35 acagcaataagtcaggaaattaatgatttagaaaaccaactcacgcaacttgaaaaagaa
 tataatcttgacttcgatttagaatatcaacaacaagttcgtgaacaatggcgatcatgct
 aaaaaataaaaaaataacttgaagaaaaacatcaatatatcaatcaatcattaacgact
 gcaaatgagcgatttagatagtttaaaacatagcattattgaaataaaaaaagagttacgt
 ttatcagaaaaaactttctgatgaattagtggttgaaagtatctcaaccattggtcaaatt
 40 aaagcgcatgataaatacattattgatttaaatcaacaacgcaataatctgctaaaagat
 atcaatcacttttatgaacgtgcacaatctgtaactgaaccacatttaaaactatttaaat
 cagatgtctttcttccatgatgtgaaacagtggttaaaaaatgcagaagaacaaaatgag
 gcttggaataaaaaatcaaactgaaacgcaattactcaataatgaattaaagcaattgaag
 tcacgcttaagtgaacgaatcaaatgattaagcaattatttgattatggtgatgtagat
 45 aatgaagaagattattatacacatcatcatcattttgaaacatatcaaagtgatttaaat
 cgatttaattgatttaaatcaatatatttagaaaatcaaaattacacttatgaaatgagttcg
 caattaagttagaaaaactactgctcaactagaagaagaagatcatagattggctaaacaa
 gttgacgattacaatgatcaatttttagaaatgcaagcagaagttagtgatttaaatgct
 cagattaatcatatggaacagatagaacttttagcacaattaagacatgaatattatagc
 50 ttaaaaaatagacttaacgatatttgctaaggattgggcaagcttaagttatatgcaagct
 ttagtggagaacatatcaagcaataaaaagataagcgtctaccacaagtgattaatgaa
 gctgtatctatttttaaaaatttaacaaatgggtacttacaatatgattcattataactgaa
 aatcataaaaatacatgtaaagcatttcaacggacaagttttgagccagttgagttgagt
 caatctacaaaagaattattatatgtggctttacgtattagttctattaaagatttaaaa
 55 ccgtattatccattcccagtgattgtagatgatgcattttgttcatttttgataaatatcgt
 aaagaacgtatggtgaaatatttgagagaactatcagaacattatcaataactttatttt
 acttgtaaaaaagatcatgtcataccggctaagaagatttaacttttaataaattacag
 gaaggcggaaaaaatga

Sequence 1500

MHEQKQKEVALHDQTQEWKRLEQSLNIEPINFPEKGIDRYETAKSHKQSLERDKSLREER
 LSILNKEAESINPVDQKYIDSFNSLYQQETEIKQKEFELRSIEKDIADKQRELEALQSN
 GWQEVFYDSTDTEAMKSHMSDLVLGKQEQIAYINQLERGLEENKIERNNSNSNEINQVENE
 5 LVPDETFEKKKEYTQQVLELHEKENLYEKLKETFEEETQKNKRQKFLRIGFIVLTILSA
 ALSIFSFFTANLIFGIIIFALLTVIFVVGIIIFSRKAVDYSTAISQEINDLENQLTQLEKE
 YNLDFDLEYQQQVREQWRHAKKNKKILEEKHQYINQSLTTANERLDSLKHSIIIEIKKELR
 LSEKLSDELVESISTIGQIKAHDKYIIDLNQQRNLLKDINH FYERAQSVTEPHLKLFN
 QMSFFHDVKQWLKNAEEQNEAWNKNQTETQLLNELKQLKSRLSETNQMIKQLFDYVDVD
 10 NEEDYYTHHHHFETYQSDLNRFNDLNQYLENQNYTYEMSSQLSEKTTAQLLEEDHRLAKQ
 VDDYNDQFLEMQAEVSDLNQINHMETDRTLAQLRHEYYSKLNRLNDIAKD WASLSYMQA
 LVEEHKQIKDKRLPQVINEAVSIFKNLTNGTYNMIHYTENHKIHKHSNGQVFEPVELS
 QSTKELLYVALRISLIKVLKPYPPFPVIVDDAFVHFDKYRKERMLKYLRELSEHYQILYF
 TCTKDHVIPAKEVLTNLKLQEGGKK*

15

Sequence 1501

Contig_0612_pos_13387_13890,
 is similar to (with p-value 1.0e-81)
 >gp:gp|U21636|SAU21636_2 Staphylococcus aureus cmp-binding-f
 20 actor 1 (cbf1) and ORF X genes, complete cds. NID: g710420.
 gtggatcatttttcttgatccatcggtgcaactcaaggtgttacagctcagggtaaaagat
 tacatgacactatcttctgcaagataaaaagtggtgatattgaagctaaattatggactgct
 acgaaagatgatatgcaaactttaaaaccagaaacaatagttcatgtcaaaggtgatatc
 atcaattatcggtgacgcaaacagatgaaaatacatcaaatacgtcttgacaaagctgaa
 25 gacaaagtgtaactaaagactttgttgacggtgcgccaatgtcacctacagaaatacaa
 gaggaattatcgcatctttatgttagatattgaaaatgctaacttacaacgcattactaga
 catttaattaaaaagtatcaagatcggtttttcacttatccagcagctagttctcatcat
 cataatttcgcgagtggtgattgagttatcatgttttaacaatgttgcgtagcaaaatct
 gtatgtgatatttatcctctgtga

30

Sequence 1502

VDHFFLIHRATQGVTAQGKDYMTLFLQDKSGDIEAKLWTATKDDMQTLKPETIVHVKGDI
 INYRGRKQMKIHQIRLAQAEDKVSTKDFVDGAPMSPTEIQEELSHFMLDIENANLQRITR
 35 HLIKKYQDRFFTYPAASSHHHNFASGLSYHVLTMRLIAKSVCDIYPL*

Sequence 1503

Contig_0612_pos_16746_17183,
 is similar to (with p-value 4.0e-52)
 >pir:pir|JC2527|JC2527 alkaline shock protein - Staphylococc
 40 us aureus >gp:gp|S76213|S76213_1 asp23=alkaline shock protei
 n 23 {methicillin resistant} [Staphylococcus aureus, 912, Ge
 nomic, 1360 nt]. NID: g894288.
 atgctcaagtgcttttttgatatacatcaagagattgattattttaaaccttctctttca
 ttgttttctttgttatctttttcattttttgttgccattctttttgtcattacgtcg
 45 tcaacttgcatgttaactcaacaacttctaaaccagtaatatattttacttggtcttta
 actaagtctgtcactttacggaataatttaggtgcagattcaccatattctaaaataact
 tttaaactctacagcagcttggttttctccaacttctacagatacgccgtgtagttacattg
 ttaccgtttgagaaagcgtagtaagctatctgtgaagccacctttcatgtctaaaatt
 cctttaacttcacgtgctgcaatacctgcaattttttcaactacttcatctgagaaagtt
 50 aatttggttttcaaattga

Sequence 1504

MLKCFDIDHQEIDYFKPSLSLFSLLSFSFFCCHSFFVITSSTCMLTSTTSKPVIIYFTCSL
 TKSVTLRKILGADSPYSKITFKSTAACFSPTSTDTFVVTLPLPEKALVKLSVKPPPFMSKI
 55 PLTSRAAIPAIFSTTSSEKVNLFNS*

Sequence 1505

Contig_0612_pos_17095_16784,
 is similar to (with p-value 2.0e-36)

>pir:pir|JC2527|JC2527 alkaline shock protein - Staphylococcus aureus >gp:gp|S76213|S76213_1 asp23=alkaline shock protein 23 {methicillin resistant}. [Staphylococcus aureus, 912, Genomic, 1360 nt]. NID: g894288.

5 atgaaaggtggcttcacagatagctttactaacgctttctcaaacggtaacaatgtaact
acagcgctatctgtagaagttggagaaaaacaagctgctgtagatttaaaagtattttta
gaatatggtgaatctgcacctaataatccccgtaaagtgacagacttagttaagaacaa
gtaaaatatattactggttagaagttggtgaagttaacatgcaagttgacgacgtaatg
10 acaaaaaagaatggcaacaaaaaatgaaaaagataacaaagaaaacaatgaaagagaa
ggtttaaaataa

Sequence 1506
MKGGFTDSFTNAFSGNNVTTGVSVEVGEKQAAVDLKVILEYGESAPKIFRKVTDLVKEQ
VKYITGLEVVEVNMQVDDVMTKKEWQQKNEKDNKENNEREGLK*

15 Sequence 1507
Contig_0612_pos_15967_15164,
is similar to (with p-value 1.0e-73)
>gp:gp|Z79580|BS168NPRB_5 B.subtilis nprB gene. NID: g162092
20 1. >gp:gp|Z99109|BSUB0006_190 Bacillus subtilis complete genome (section 6 of 21): from 999501 to 1209940. NID: g2633260
. >gp:gp|Y09476|BSY09476_54 B.subtilis 54kb genomic DNA fragment. NID: g2145361:
atgcaaccttattttaattgtctagatctagatggtacattattaaatgacaataaagaa
25 atctcaccttacactaaacaagtatttaaccgaattacaacaatgtggacactacgttatg
attgctacttggagaccctatcgcgcaagccagatgtattatcatgaactaaatatgagc
acacctgttgtaactttaatggagcatttgtacatcatccaaaagcaaacgattttaaa
gtgatacatgaagtacttgatgtggaaatttctaaaaatattattacagcacttcaacaa
tctcatattacaaatatcattgctgaagtaaaagactatgtctttataaatagttatgat
30 tcaagactttacgaaggtttttcaatgggaaatcctaaaattcaaacaggtaatctactt
gaaaatcttaatgaagcacctacgtcattacttgttgaagcagaagaagaaaatattcct
gaaattaaagatatgttaacacatttttatgcagaaaatattgaacatcgtcgttggggc
gcaccgtttccagtaatagaattgtgaagcgtgggattaaacaaagcacgtggaatcaag
catgttcaaaactatttaaacatcgccgacgatcatatcattgcgtttggtgatgaggac
35 aatgatatagaatgataaagtttgcgacccatggcattgcaatggccaatggcttgaaa
gatttaaaggaaatagcaaatgagactacgtatagtaataatgaagacggaataggctcgt
tatttaaatgacttttggtattaa

Sequence 1508
40 MQPYLICLDLDGTLNNDNKEISPYTKQVLTELQCCGHYVMIATGRPYRASQMYHELNMS
TPVVNFNGAFVHHPKANDFKVIHEVLDVEISKNIITALQQSHITNIIAEVKDYVFINSYD
SRLYEGFSMGNPKIQTGNLLENLNEAPTSLLVEAEEENIPEIKDMLTHFYAENIEHRRWG
APFPVIEIVKRGINKARGIKHVQNYLNIADDHIIAFGDEDNDIEMIKFATHGIAMANGLK
DLKEIANETTYSNNEGIGRYLNDFWY*

45 Sequence 1509
Contig_0612_pos_8705_8037,
is similar to (with p-value 9.0e-86)
>gp:gp|AF076683|AF076683_5 Staphylococcus aureus oligopeptide transporter putative substrate binding domain (opp-1A), oligopeptide transporter putative membrane permease domain (opp-1B), oligopeptide transporter putative membrane permease domain (opp-1C), oligopeptide transporter putative ATPase domain (opp-1D), and oligopeptide transporter putative ATPase domain (opp-1F) genes, complete cds; and unknown gene. NID: g3800817.
50 gtgtcatttgattgccccactggtgcatcaatagcaattattggagaaagtggagtgga
aagtctacgttgatcgctatgatttttaggactagaaaaaccagatcaaggacaagtgacg
ttggacggtcagcctgttcatttaaaaaaagtgagacgtcatcgaattgctgcggttttt

caagactatacttcttcttgcacccccatactgtcaaagatattttggttgagta
 atgaatcagtgatgatgacatcaaaggagaatatggaagagtgataacagcggtacta
 cgtgaagtaggggttaaaatctgattgtttgtattgttatccacatatgctttcaggtgga
 gaagcacagcgtgtagctattgcacgtgcatatgtatgcaaccagattatatattatt
 5 gatgaggcgattagttcattagatatgtctatgcaaacacaaatattagatttattgaaa
 aggttacgtcactcacatcagctgagttatatttttattactcatgatatacaagctgcc
 acgtatatatgtgatgacttgcttatttttaaaaatggctgtatcgaagctaggacatct
 ataagtgaattgcataggcaacaaaatggttatacaagagaactgattgataaacaacta
 tcaatctaa

10

Sequence 1510

VSFDCPTGASIAIIGESGSGKSTLSRMILGLEKPDQGVTLTGQPVHLKKVRRHRRIA
 QDYTSSLHPFHTVKDILFEVMNQCRCTSKENMEEYVTALLREVGLKSDCLYCYPHMLSGG
 EAQRVAIARAICMPDYILFDEAISSLDMSMQTQILDLLKRLRHSQLSYIFITHDIQAA
 15 TYICDDLILFKNGCIEARTSISELHRQNGYTRELIDKQLSI*

Sequence 1511

Contig_0612_pos_7974_6835,

is similar to (with p-value 0.0e+00)

20 >gp:gp|AF076683|AF076683_6 Staphylococcus aureus oligopeptid
 e transporter putative substrate binding domain (opp-1A), ol
 igopeptide transporter putative membrane permease domain (op
 p-1B), oligopeptide transporter putative membrane permease d
 omain (opp-1C), oligopeptide transporter putative ATPase dom
 25 ain (opp-1D), and oligopeptide transporter putative ATPase d
 omain (opp-1F) genes, complete cds; and unknown gene. NID: g
 3800817.

atgttttttagtgcaatgccatactcaatgttttcatacctctaagaggacatgacttg
 gaggcgacgaataaccgtaattggaattgtaatgggagcttacatgctaacggcaatgcta
 30 tttcgcccttgggctgggtcaaattattgcacgtgtaggaccgattaaagtattgcgtatt
 atattattgattaatgctatggcactgggtattatatgggtttacaggacttgaaggttat
 ttggttgacgtatcatgcaaggtgtatgtacggcattcttctcaatgtctttacaattg
 ggtattatagatgctttacctgaaaaatatcggtcagaaggtgtatctctctattcattg
 ttttcaacaattcccaatttattaggaccattaattgcagttgggatttggcacgtggaa
 35 aatatgaccatatattgctattgttatgatttttattgcagtaacaacaaccttatttgg
 tatagaactacttttgcaaatacacaaaagaggttagcaccacaaagaagaagtcctacct
 tttaatgcaatgactgtatatgttcaatttttaaaaataaagcactcttctgcagtggt
 atgattatgattttgtcatctatcggtttggtgcgatgagtactttataaccattatat
 acagtttagggaagggtttcgcaatgcaggtattttccttacaattcaagccattacagta
 40 gtgatagctagatttttattacgtgaagtatgtaccatctgatggtttatggcatcaccgg
 tttatgatgatgtgtcttaacgttactgatgattgcttcaatcattgtagcttttgacca
 caaatattgagtatatattgtatatataagtgcgaatctttattggaataacacaagcgctc
 gtttatcctacattgacaacgtatttaagttttgtcttaccacaaagataggacgtaatatg
 ttattaggattgtttatagcatgtgcagatttagggatttcactaggaggtgtgctaag
 45 gggccaatatcagatacggtaggatttaaatggatgtatattttatgcgctttattggt
 actattgcaatgatactaagtaaaattagacaaggacaaagtgtttctaaagcttcatag

Sequence 1512

50 MFFSANAILNVFIPLRGHDLEATNTVIGIVMGAYMLTAMLFPRWAGQIIARVGPIKVLRI
 ILLINAMALVLYGFTGLEGYLVARIMQGVCTAFFSMSLQLGIIDALPEKYRSEGVSLYSL
 FSTIPNLLGPLIAVGIWHVENMTIFAIVMIFIAVTTTLFGYRTTFANTQKEVAPKEEVLP
 FNAMTVYVQFFKNKALFCSGMIMILSSIVFGAMSTFIPLYTVREGFANAGIFLTIAITV
 VIARFYLRKYVPSDGLWHRFMMIVLTLLMIASIIIVAFGPQILSIFVYISAFIFIGITQAL
 55 VYPTLT'TYLSFVLPKIGRNMLLGLFIACADLGISLGGVLMGPISD'TVGFKWMIYLCALLV
 TIAMILSKIRQGQSVSKAS*

Sequence 1513

Contig_0612_pos_3672_3067,

is similar to (with p-value 2.0e-47)
 >gp:gp|AF051356|AF051356_5 Streptococcus mutans YtqB (ytqB)
 gene, partial cds; ABC transporter (abcX), putative permease
 (perM), putative hemolysin (hlyX), pyruvate-formate lyase a
 ctivating enzyme (pflC), D-alanine-D-alanyl carrier protein
 5 ligase (dltA), integral membrane protein (dltB), D-alanyl ca
 rrier protein (dltC), extramembranal protein (dltD), and put
 ative exopolyphosphatase (ppx1) genes, complete cds; and unk
 nown gene. NID: g2952523. >gp:gp|AB018417|AB018417_2 Strepto
 10 coccus mutans genes for PFL-activating enzyme and PFLAE-5'OR
 F, partial and complete cds. NID: g3986292.
 gtgacggttgatgaaatggtaaatgaaatcttaccgtacaaaccttactttgaagcttca
 ggtggtggggaatacagtcagtggtggcggaaccattactacaaatgcctttcttgagcaa
 ttattcaaagaattaaaagcgaatggtgttcacacatgcattgatacttctgcgggatgt
 15 gtgaatgatacaccagcatttaaatcgtcattttgatgaattgcaaaagcatacagattta
 atcttattagatattaaacatattgataatgataagcacatcaaattaacaggcaaacct
 aacacacatatatttaaagtgttcacgtaaattatctgatatgaaacaacctgtttggatt
 agcatgttttagtacctggtatttcggatgataaagaagatttgataaaactaggagaa
 ttattaattctttagataacgttgaaaagttgaaatcttaccatatcatcaactcgg
 20 gtgcataagtggaaaaatttaggcattcccttatcaactcgaaaatgttgaaccatctgac
 gatgaagcgggttaaagaagcttatcgctatgttaactttaatggcaaaataccgtaaca
 ttatag

Sequence 1514
 25 VTVDEMVNEILPYKPYFEASGGGVTVSGGEPLLQMPFLEQLFKELKANGVHTCIDTSAGC
 VNDTPAFNRHFDELQKHTDLILLDIKHIDNDKHIKLTGKPNTHILKFARKLSDMKQPVWI
 RHVLVPGISDDKEDLIKLGFINSLDNVEKFEILPYHQLGVHKWKNLGIPLYQLENVEPSD
 DEAVKEAYRYVNFNGKIPVTL*

Sequence 1515
 Contig_0612_pos_1364_3,
 is similar to (with p-value 3.0e-98)
 >sp:sp|P54104|BRNQ_LACDL BRANCHED CHAIN AMINO ACID TRANSPORT
 SYSTEM CARRIER PROTEIN. >pir:pir|S60180|S60180 branched-cha
 35 in amino acid carrier brnQ - Lactobacillus delbrueckii >gp:g
 p|Z48676|LDBRNQGN_1 L.delbrueckii brnQ gene for branched-cha
 in amino acid carrier. NID: g732812.
 atgatgaaaaataaattaacattaaaaagagaatctatttatcggctcaatgctgtttggt
 ctttttttgggtgctggaaatctcatttttccaattcacttaggtcaaaactgcgggggca
 40 aatgtatggaccgcaaatatttaggatttcttatcacggctatcggactaccttttttagga
 attatagcgataggtgtatctaaaacaacggggtctttgaaatttcctcaaggataagt
 aaaatatatggttatttggttcacattggcttgatctgttataggtccgttttttgcg
 ttgccaaagacttgcgacgacgtcatttgaaatagcatttccaccatttattcatctggt
 acggcccaagcgttgttgccatttttttagtattttattcttcggagtagcgtggttattt
 45 tcgcgtaaaccttctaaaatatttagactatattggaaaattcttaaatccgggtctttctc
 atcttgcttggaattgttgggttgcttgcatattatccgtcctatgggtggaattagtcat
 gcgccagtaagtgtgattatagcaatagcgtgttactcaaagggtttatcgatggatat
 aatacattagacgctttggcatcattagcatttgggtattatcattgttactacaattaaa
 aagttggggattactaatccgaatacaatcgctaagaaactttaaaatcaggtacgatt
 50 agtattatagctatgggcgttatttatactttattagctttaatgggtacgatgagttta
 ggtcgttttaaagtaagtgaatggtggtattgcgcttgctcagattgcacaacattat
 ttaggggattacggaattattatttggcactaatcatcattgtggcatgtctgaaaaca
 gcaataggattgatcacagccttttcggaacatttacagagttattccctaataatctaac
 tatctttgggtgactactgggtgagtatattagcttgatatatttgctaagttaggttta
 55 acaaaaattattatgtattcaacaccagtggtgatgttcatttatccttttagcgattact
 ttaattttattagcattacttagtccattatttaaacattctaaaattgtctatcgattt
 acaacattatttacaatgggtggcggtattttagatggtgtgaaagcaagtccagagttc
 tttgttaatacaaaaatttgcacaaacaatcattggatttgggtgaaaattatctccattc
 ttaacattgggtatgggatggattgttccagcacttattgggttcattattgggtattatt

gtatactttatgactgctaaaaaatcgctccacgtacaataa

Sequence 1516

MMKNKLTLENLFIGSMLFGLFFGAGNLIFPIHLGQTAGANVWTANLGLLITAIGLPFLG
 5 IIAIGVSKTNGVFEISSRISKIYGYLFTIGLYLVIGPFFALPRLATTSFEIAFSPFISSG
 TAQALLPIFSILFFGVAWLFSRKPSKILDYIGKFLNPVFLILLGIVVVLAFIRPMGGISH
 APVSADYSNSVLLKGFIDGYNTLDALASLAFGIIIVTTIKKLGITNPNTIAKETLKSGTI
 SIIAMGVIIYTLALMGTMSLGRFKVSENGGIALAQIAQHLYGDIYGIILSLIIIVACLKT
 10 AIGLITAFSETFTELFPKSNYLWLATGVLSILACIFANVGLTKIIMYSTPVLMMFIYPLAIT
 LILLALLSPLFKHKSIVYRFTTLFTMVAAFVDGVKASPEFFVNTKFAQTIIGFGENYLPF
 FNIGMGWIVPALIGFIIGIIVYFMTAKKSSHVQ*

Sequence 1517

Contig_0613_pos_3835_4197,
 15 putative peptide of unknown function
 atgaatgttgagtaatttcaaagttccaaaagttagattaggaatagaacttatagtcaa
 agcgagctacaagactataggaaagccaatacacaaaggtataaccaagaggttagacac
 aataggcacaatagagagtatacagcgttctacaacagtacacagtggtcgtaagttgcgt
 aaacaagtattattacgtgataactacttctgtcaacattgtttaagtaaagggaatagt
 20 aatgacaaagattttgattgttcaccataagattgaattaaaacgggactggtcgaaaaga
 ctggatatggataatttagaggcagtggtgttttagctgccataataaaattcacggtgga
 taa

Sequence 1518

25 MNLSNFKVPKVRNLGNRTYSQSELQDYRKANTQRYNQEVNRHNRHNREYTAFYNSTQWRKLR
 KQVLLRDNYLCQHCLSKGIVNDKDLIVHHKIELKRDWSKRLDMNLEAVCFSCHNKIHGG
 *

Sequence 1519

30 Contig_0613_pos_8339_7815,
 is similar to (with p-value 1.0e-49)
 >sp:P08064|DHSC_BACSU SUCCINATE DEHYDROGENASE CYTOCHROME
 B-558 SUBUNIT. >pir:pir|A29843|DEBSSC succinate dehydrogenas
 e (EC 1.3.99.1) cytochrome b558 - Bacillus subtilis >gp:gp|M
 35 13470|BACSDHAB_1 B.subtilis succinate dehydrogenase complex
 encoding cytochrome b-558 subunit, complete cds, and flavopr
 otein subunit, 5' end. NID: g143524. >gp:gp|Z99118|BSUB0015_
 110 Bacillus subtilis complete genome (section 15 of 21): fr
 om 2795131 to 3013540. NID: g2635200. >gp:gp|Z75208|BSZ75208
 40 _57 B.subtilis genomic sequence 89009bp. NID: g1769994.
 atggtaaacatcaagcaacgcaaggtgctgaagcttttaatagagcttcaggatttatg
 gaatctttaccattccttattgtgatggaatttatacttatttatataccattgttatac
 catggtttgttcggtttacacatcgcatcactgctaaggagaacatcgggcattactca
 ttatttagaaactggatgtttttcttccaacgtgtaagtggatatttagcattgttttt
 45 attgcaatgcacttatggcaaacacggtttgcaaaaagctttttatggtaaatctgtggac
 tataatctaataatgcataaacattacaacatccgttatgggcaatctttacattattgt
 gtcattgctgtgtttttccattttgctaataatggtttatggatctttgtgtaacatggggc
 tttttacaatctaaaaaatcacaacgtgtttttacttgatttctactcatagttatttta
 gtgatttcttatattgtgtgtgcagccgttattgcgtttatataa

Sequence 1520

MVNHQATQGAFAFNRSAGFMESLPFLIVMEFILIYIPLLYHGLFGLHIAFTAKENIGHYS
 LFRNWMFFQVRVSGILAFVFIAMHLWQTRLQKAFYGKSVVDYINLMHETLQHPLWAIIFYIIC
 VIAVVEHFANGLWSFCVTWGFLQSKKSQRVFTWISLIVFLVISYIGVAAVIAFI*
 55

Sequence 1521

Contig_0613_pos_7614_6403,
 is similar to (with p-value 0.0e+00)
 >pir:pir|A27763|A27763 succinate dehydrogenase (EC 1.3.99.1)

flavoprotein - *Bacillus subtilis*

atgtcaacaattaaagcggcagaacaaggtgcacatgtagatttattttccattgtaccg
 gtaaagcgttcgcactctgtttgtgcacaaggtggcataaatggtgctgttaatactaaa
 ggtgagggagattcaccgtggattcactttgatgatactgtttatggtggagacttcctg
 5 gctaatacaaccaccagtc aaagcaatggctgatgctgcacctaaaatcatccatctgtta
 gatcgtatgggggttatgtttaacagaacgaaagaaggcttatttagacttttagacgtttt
 ggcggtacactacatcatagaacagctttttgctggcgcaacgacaggtcaacaattgctt
 tatgcattagatgagcaagttcgttcatttgaggtagatggtttagtaactaaatacgaa
 ggatgggaatttctaggtattgttaaagacgaagaagatgctgcaagaggtattgttgct
 10 caaaatatgacaacatcagaattcaatcattcggttcagatgctgtcatcatggcaaca
 ggtggctcctggattattcttggtaaaacgacgaattcaatgattaatacaggttcagcg
 gcgtcaatcgtttatcagcaaggtgcgatttatgcaaatggtgaattcatccaaatacat
 ccgactgcgattcctggagatgacaaattacgtcttatgagtgaatcagctcgtggtgaa
 ggtggacgtatttggacgtataaagatggtaaaccttggacttcttagaagaaaaatat
 15 ccagactatggtaacttggttccacgtgataatagcgacacgtgaaattttcgatgtttgt
 attaaccaaaaggttaggtatcaatggagaaaacatggtataccttgatttatctcataaa
 gatccacacgaattggatgttaaatagggtgattattgaaatttatgaaaaattcaca
 ggtgatgattccacgtaaagttccaatgaaaatctcccagcagtgcaattattcaatgggt
 ggtttatacgtagactatgatcaaatgactaatatcaaaggggtatttgcagctggagaa
 20 tgtgatttctcacaacatggtggtaaccgtttaggtgccaatcttttactttcagctatt
 tataacataagcaaatatggtagccatgagtggtatctatttttctcggttcattctctt
 ttcaatcattaa

Sequence 1522

25 MSTIKAAEQGAHVOLF SIVPVKRSHSVCAQGGINGAVNTKGE GDS PWIHFDDTVYGGDFL
 ANQPPVKAMADAAPKIIHLDRMGVMFNRTKEGLLD FRRFGGTLHHR TAFAGATTGQQLL
 YALDEQVRSFEVDGLVTKEYEGWEFLGIVKDEEDAARGIVAQNMTTSEIQSFGSDAVIMAT
 GGP GII FGKTTNSMINTGSAASIVYQQGAIYANGEFIQIHPTAIPGDDKLRLMSESARGE
 GGRIWTKDGKWPWFLEEKYPDYGNLVP RDIATREIFDVCINQKLGINGENMVYLDLSHK
 30 DPHELDVKLGGIIIEIYEKFTGDDPRKVP MKIFPAVHYSMGGLYVDYDQMTNIKGLFAAGE
 CDFSQHGGRNLGANSLLSAIYNISKYGSHEWIYFFSVHLLFNH*

Sequence 1523

Contig_0614_pos_6876_6298,
 35 is similar to (with p-value 7.0e-19)
 >gp:gp|U40604|LMU40604_2 *Listeria monocytogenes* ClpC ATPase
 (mec) gene, complete cds. NID: g1314293.
 gtgaggtgtttaaaattgctttgtgaaaattgccatttttaataagcgggaagttaaactt
 actgttaaaggatatagatagtagcatgaaaaatgggtatgttcagtatgtgccaagga
 40 gaaaacccttggttacattctaacgatgataatacgtatcatacacaccaagacgatata
 gaagaagcattttagtgaaacagatacttcaacaccttgctgcaaaacatggtattaat
 tttcatgagatggcatttaaagaagaaaaaaatgcccaacgtgtcagatgacacttaag
 gatattgcacatgttggaagcttgggtgtgctgattgttatgctacgtttaaagaagac
 atcattgatatagttcaacgtgttcaaggtggtcaatttgaacatgtagaaaaacacca
 45 caatcatcgtataagaacttgcaataaaaaagcaaattgaagaaaaatcaaaatatcta
 aataaattgatagatgggtcaagagtttgaagaggcagcgattgttcgtgatgaaattaa
 gctttaaaagtgagagcgaggtgtctcatgatgagtaa

Sequence 1524

50 VRCLKLLCENCHFNEAEVKLT VKGIDSTHEKWVCSVCAQGENPWLHSNDNTYIITHQDDI
 EEAFVVKQILQH LAAKHGINFHEMAFKEKKCPTCQMTLKDIAHVGLKGCADCYATFKED
 IIDIVQRVQGGQFEHV GKT PQSSYKKLAIKKQIEEKS KYLNKLIDGQEFEEAAIVRDEIK
 ALKSESEVSHDE*

55 Sequence 1525

Contig_0614_pos_6266_5301,
 is similar to (with p-value 3.0e-70)
 >sp:sp|P37570|YACI_BACSU HYPOTHETICAL 41.1 KD PROTEIN IN LYS
 S-MECB INTERGENIC REGION (ORFX). >gp:gp|D26185|BAC180K_147 B

. subtilis DNA, 180 kilobase region of replication origin. NID: g467326. >gp:gp|Z99104|BSUB0001_85 Bacillus subtilis complete genome (section 1 of 21): from 1 to 213080. NID: g2632267.

5 atgtctgaggagacacctgttattattttcttcagaattcgattagctagaaatcttgaa
aaccatgtccaccacttatgttcccttcagagcaagaaggatatcgagtataaatgaa
gttcaagatgagctttccaacttaactttaaatcgattagatacgatggatcaacaaagt
aaaatgaaattggttgcgaacatcttgtgagtcctgaactagtgaacaacctgcttca
gcagtaattgttaaatgatgatgaatcggttaagtgttatgataaacgaagaagatcatata
10 cgaatacaggctctaggaactgatttatcgctaaaggatttatatcaacgcgcttctaaa
attgatgatgaattagataaaagcgttagacattagttatgatgagcatttaggatattta
actacctgtcctactaataatttggtacaggaatgcgtgcaagtgtgatgttacatttacct
ggactctccattatgaaaagaatgaacagaattgcacaaacaattaatcgttttggattt
acaattcgaggtatatacggagaagggtcacaagtatatgggtcacatttatcagggtttca
15 aaccaacttacactagggaacagaagaagacattatcgataacttaactgaagttgta
aatcaaatataaatgaagaaaagcaataagagaaagacttgataaacacaattctgta
gagacactggatagagtttatcgatcattaggtgtactacaaaacagtagaattatttct
atggaagaagcctcatatcgtttgagcgaagtgaactaggtattgattgaattatatt
ttgcttgaaaattttaaatgaattaatggtagcaatacagtcaccatttttaata
20 gatgacgatgataatagaacagtaaatgaaaaagagctgatttattaagagaacatata
aaatag

Sequence 1526

MSEETPVISSRIRLARNLENHVVHPLMFPEQEGYRVINEVQDALSNTLNRLDTMDQQS
25 KMKLVAKHLVSPKQFASAVMLNDDDESVMINEEDHIRIQALGTDLSLKDLYQRASK
IDDELDKALDISYDEHLGYLTTCPTNIGTGMRSVMLHPLGLSIMKRMNRIAQTINRFGF
TIRGIYEGESQVYGHYQVSNQLTLGKTEEDIIDNLTEVVNQIINEEKQIRERLDKHNSV
ETLDRVYRSLGVLQNSRIISMEEASYRLSEVKLGIDLNYILLENFKFNELMVAIQSPFLI
DDDDNRTVNEKRADLLREHIK*

30

Sequence 1527

Contig_0614_pos_5287_2834,

is similar to (with p-value 0.0e+00)

>sp:sp|P37571|MECB_BACSU NEGATIVE REGULATOR OF GENETIC COMPE
35 TENCE MECB. >gp:gp|D26185|BAC180K_148 B. subtilis DNA, 180 k
ilobase region of replication origin. NID: g467326. >gp:gp|U
02604|BSU02604_2 Bacillus subtilis Marburg 168 ClpC adenosin
e triphosphatase (mecB) gene, complete cds, orfX and orfY, p
artial cds. NID: g442358. >gp:gp|Z99104|BSUB0001_86 Bacillus
40 subtilis complete genome (section 1 of 21): from 1 to 21308
0. NID: g2632267.

atgttatttggtagattgacagagcgtgcacaacgtgtgttggcacatgcacaagaggaa
gcaattcgtttgaaccatttctaataattggaacagaacatcttttgcttggttaaatgaaa
gagccagaaggtatagcagcaaaggtattagtaagttttaataattactgaagataaagtc
45 atcgaagaagttgaaaaacttatcggtcacgggtcaagagcaaatgggcacactacattat
acaccgagagcaaaaaagtaattgaactgtctatggatgaagctcgaaagctacatcat
aactttgtaggaacagagcatatactattaggtttaattagagaaaatgaaggtgttgca
gcacgtgtatttgcacacctagatttaaatattactaaagcacgtgcccaggttgtaaaa
gcttttaggaagtcagaaatgagtaataaaaaatgcgcaagctaataagtctaataacacg
50 cctacttttagatggattagctagagatttaactgttatttgctaaagatggaacgtagat
ccagtcgtaggacgagataaagaaattactcgtgtaattgaagttttaagtcgctcgtact
aaaaataatcctgtgctaattgggtgaacccgggtgttggtaaaacagcaattgctgaaggg
cttgcgcaagcaattgttaaaaatgaagtaccagaaactttaaaagacaaacgtgtaatg
tcattgatattgggtacagtcgtagctggcactaaatatcggtggtgaatttgaagaaaga
55 ttgaaaaaagttatggaggaaatccatcaagctggtaattgttattctatttatcgatgaa
cttcatacttttagttggcgctgggtggcgagaggaagcaattgatgcatctaataatttta
aaacctgcttttagctcgtggagaattgcaatgtataggtgccacaacatttagatgaatat
cgtaaaaaatatagaaaaagacgctgcattagaacgtcgttttcaaccaattcaagtggt
gaacctacagttgaagacacgattgaaatcttaaaaggattacgtgaccgttatgaggct

catcacagaattaatatctcagatgaagcctttagaagcggctgctaaattgagtgatcgc
 tatgtttcagatcggttcttgcagataaagccattgacttaattgatgaggcaagttca
 aaagttagacttaaaagtcatacaacgccaagtaatttaaaagagattgaacaagaaatt
 gataaagtaaaaaatgaaaaagatgctgcagttcatgctcaagaatttgaaaatgccgct
 5 aatttaagagataagcaatctaaacttgaaaagcaatatgaagatgctaaaaatgaatgg
 aaaaaatgcacaaggtgggttagatactgccttatctgaagaaaatatcgctgaagtaata
 gctgggtggacaggtattcctttaactaaaattaatgaaactgaatcagatcgtttattg
 aatcttgaagatacacttcataaacgtgtcattggacaaaacgatgctgtcaattcaatt
 10 agtaaaagctggttagaagagctcgctgctggctttaaagatccaaaacgtccaatcggtagt
 tttatcttcttaggacctacaggtgtgggtaaaactgaattggctcgctgctttagctgaa
 tctatgtttgggaagacgatgcaatgattcgctgagatagagtgaatttatggagaaa
 catgctgtcagtcgattagttgggtgcacctccaggatatgtaggacatgatgacggcggt
 caattgactgaaaaagttagacgtaaacactactctgtgattttatttgatgaaattgag
 aaagcacatcctgacgtatttaatatcttctacaagtttagatgatggctcatttaaca
 15 gatactaaaggtcgctactgtggacttccgtaatactgtgattattatgacttctaattgtg
 ggagctcaagaattacaggaccaacgcttctgctgggtttggagggtgcttcagaaggtagt
 gactacgaaactgtcagaaaaacaatgatgaaagaattaaaaaattcattccgaccagaa
 ttcttaaacctggttgatgacattattgtcttccacaaacttacaaaagatgaattaaaa
 gaaattgttacaatgatggttaataaaacttactcaccgtcttccagagcaaaatattaat
 20 attgtgttactgataaagcgaaagaaaaaattgcagaagaaggatagatcctgaatat
 ggtgctagaccactcattagagcaattcaaaaaacggttgaagataatttaagcgaattg
 attttagatggaaataaaattgaaggtaaagaagtaacaattgatcatgatggttaaagaa
 ttaagtatgatatttatgaaattacagctaaaaaagaacaacagaatcataa

25 Sequence 1528

MLFGRLTERAQRVLHAHQEEAIRLNHSNIGTEHLLGLMKEPEGIAAKVLVSFNITEDKV
 IEEVEKLIGHGQEQMGLHYTPRAKKVIELSMDEARKLHHNFVGTGEHILLGLIRENEGVA
 ARVFANLDLNTKARAQVVKALGSPMSNKNQAANKSNNTPTLDGLARDLTVIAKDGTL
 30 PVVGRDKEITRVIEVLSRRTKNNPVLIGEPGVGKTAIAEGLAQAIKNEVPETLKDKRVM
 SLDMGTVVAGTKYRGEFEERLKKVMEEIHQAGNVILFIDELHTLVGAGGAEGDAIDASNIL
 KPALARGELQCIGATTLDEYRKNIEKDAALERRFQPIQVDEPTVEDTIEILKGLRDRIYA
 HHRINISDEALEAAAKLSDRYVSDRFLPKAIDLIDEASSKVRKLSHTTPSNLKEIEQEI
 DKVKNKEDAAVHAQEFENAAANLRDKQSKLEKQYEDAKNEWKNAQGGLDTALSEENIAEVI
 AGWTGIPLTKINETESDRLLNLEDTLHKRVIGQNDVNSISKAVRRRARAGLKDPKRPIGS
 35 FIFLGPTGVGKTELARALAESMFEGDDAMIRVDMSEFMEKHAVSRLVGAPPGYVGHDDGG
 QLTEKVRKPYSVILFDEIEKAHPDVFNILLQVLDDGHLTDTKGRTVDFRNTVIIMTSNV
 GAQELQDQRFAGFGGASEGSDYETVRKTMKELKNSFRPEFLNRVDDIIVFHKLTKDELK
 EIVTMMVNKLTHRLSEQNINIVVTDKAKEKIAEEGYDPEYGARPLIRAIQKTVEDNLSEL
 ILDGNKIEGKEVTIDHDGKEFKYDIYEITAKKETTES*

40 Sequence 1529

Contig_0614_pos_1884_970,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P37572|RADA_BACSU DNA REPAIR PROTEIN RADA HOMOLOG (DN
 45 A REPAIR PROTEIN SMS HOMOLOG). >gp:gp|D26185|BAC180K_149 B.
 subtilis DNA, 180 kilobase region of replication origin. NID
 : g467326. >gp:gp|Z99104|BSUB0001_87 Bacillus subtilis compl
 ete genome (section 1 of 21): from 1 to 213080. NID: g263226
 7.
 50 gtgatacatcaaaactgtaaaagaagagagacctgacttacttggtgttgattcgattcaa
 acaatctatcatccggaaattagttccgcacctggatcggtatcacaagtaagagagagt
 acgcagaggttaaatgaacattgctaacaacaaatgaatattgccacatttattgtgggacac
 gtaacaaaagaaggacaaatcgccggaccaagattattggaacatatggttgatacagtt
 ctttattttgaaggagatgagcatcacgcatatcgatccttagagcagtaaaaaataga
 55 tttggttctacaaatgagatggggattttcgaaatgaagcaaagtggtataaaagggtgta
 cttaatccttctgaaatgtttttagaagaacgttctacaaatgttccgggctctacaatc
 gtccccactatggaaggaaacagaccactactcattgaagttccaagcgcttgttacacca
 acaacatttaataatcctagacgaatggctacaggtatagatcataatcgattaagttta
 cttatggcgggttctagaaaaaaggaaaactatttactccaacaacaagatgcctatatt

aaagtagcaggtggcgtcaaattaacagaacctgctggtgatttaagcattattggtgcg
 acagcttcaagttttaaagatcaagctggtgatggattagattgttttgtgggtgaagtt
 ggattaacaggtgaagtacgcagagtatctcgcatagagcaacgtgttcaagaagcgacc
 aaactagggtttaaaagagctattattccacagacaaatattggaggttgacattccca
 5 gaaggcatccaagtcgttggtgttcatcagtagcatgaagctttgaaatatgcattacat
 tcaaacacagcgataa

Sequence 1530

VIHQTVKEERPDLLEVDSIQTIYHPEISSAPGSVSQVRESTQSLMNIKQMNIATFIVGH
 10 VTKEGQIAGPRLLEHMDTVLYFEGDEHHAYRILRAVKNRFGSTNEMGIFEMKQSGLKGV
 LNPSEMFLERSTNVPGSTIVPTMEGTRPLLIEVQALVPTTFNNPRRMATGIDHNRLSL
 LMAVLEKKENYLLQQDAYIKVAGGVKLTEPAVDLSIIIVATASSFKDQAVDGLDCFVGEV
 GLTGEVRRVSRIEQVRVQEATKLGFKRAIIPQTNIGGWTFPEGIQVVGVSSEALKYALH
 SKQR*

Sequence 1531

Contig_0614_pos_0_934,
 is similar to (with p-value 1.0e-46)
 >gp:gp|U40604|LMU40604_6 *Listeria monocytogenes* ClpC ATPase
 20 (mec) gene, complete cds. NID: gl314293.
 atggaaggaggtatataaattgaatataacaaaagcaattgtttagcaatctatatcatt
 gttggtgcagcacttgggtgtataattataaccgaagttgttacagatcttggcattcat
 caccatgcggttatcactaattattatgtagatggtttcatagggatcattatattttt
 ataattattggtattgttcattaataaagtaacatatgcttttaacaatttgaacaatta
 25 atcatgagacgtagtgccgtagaataattatttgcataaattggtttaattattggttta
 tttatttcagtgtggtttcttttatcttagaaatgataggttaattccatattaaatcac
 tttgtacctatgataatcactatttttatgttatttagggtttcaatttggctctgaaa
 aaaagagatgaaatgcttatgtttttaccagagaatatggcacgttccatgtctaataat
 atacgaagagcgacacctaagattgtagatacaagtgccattatcgatggaaggatatta
 30 gatattatacgttgcggatttatcgatggtgatattgataccacaaggcgttataaat
 gaattacaggttatagcggatgctaaagatagcgtgaaacgtgaaaaaggtcaaagagga
 ttagatattttgaatcaactttatgatttagattatcctacacgcgttatacatccaact
 caatcccatagtgatatagatacattattaattaaattagcacacagtatcatgcacat
 gtgattacgactgatttttaatttaataaagtatgtcacgttcaaggaattacagcactc
 35 aacgttaatgatttatcggaagcaatcaaactaatgtacatcaaggcgaccagttaagt
 attttattaacgaagataggttaagagcCTTTTAA

Sequence 1532

MEGGYKLNITKAIIVVAIYIIIVGAALGVIIIEVVTDLGIHHHAVITNYYVDGFIGI I IFF
 40 IIFGLFINKVTYAFKQFEQLIMRRSAVEILFATIGLIIGLFISVMVSFILEMIGNSILNH
 FVPMIITIIILCYLGFQFGLKKRDEMLMFLPENMARSMNNIRRAPKIVDTSI I DGRIL
 DIIRCGFIDGDILIPQGVINELQVIADAKDSVKREKQGRGLDILNQLYDLDPTRVIHPT
 QSHSDIDTLIKLAQQYHAHVITDFNLNKKVCHVQGITALNVNDLSEAIKPNVHQGDQLS
 ILLTKIGKELFX

Sequence 1533

Contig_0615_pos_391_900,
 is similar to (with p-value 1.0e-48)
 >gp:gp|U00013|U00013_9 *Mycobacterium leprae* cosmid B1496. NI
 50 D: g466868.
 atgaaatatccaaactgtgtacttttaggtgaagggtgcaaaggtagcacattatccatt
 gcatttgcgtgtaagggtcaagttcaagatgctggtgctaaaatgattcataaagcacct
 aatacatcttcaactattgtttctaaatctatctccaaaaatgggtggtaaagtcatttat
 cgtggtatcggttcatttttgacgtaaagctaaggagcacgttcaaataatcgatgtgat
 55 acattaatttttagataatgaatcgacttcagatactatcccttataatgaagtgttcaat
 gacaatatctcattagaacatgaagctaaagtttctaaagatcagaagagcaattattc
 tatcttatgagtcgtggtatttctgaggaagaagcgacagaaatgattgttatgggattc
 attgagccgttcacaaaaagaattaccaatggaatacgcagtagaaatgaaccgtttaatt
 aagtttgaaatggaaggctcaattgggttaa

Sequence 1534

MKYPNCVLLGEGAKGSTLSIAFAGKGQVQDAGAKMIHKAPNTSSTIVSKSISKNGGKVIY
 RGIVHFGRKAKGARSNIECDTLILDNESTSDTIPYNEVFNDNISLEHEAKVSKVSEEQLF
 5 YLMSRGISEEEATEMIVMGFIEPFTKELPMEYAVEMNRLIKFEMEGLSIG*

Sequence 1535

Contig_0615_pos_2732_3145,
 putative peptide of unknown function

10 atgatgaataaagcaattaataataacatcattgataggaatcattttacaaagtttttct
 agtctattgttttttagttttcttggatattttcaattactggagcgatggatgctaatttt
 actacaacagtttaattgtgaaacaacagttcatagtgcagaagccgcacagcatgtttt
 agtggttggtatttttggttatattcatcggttgaattttttcaatcatttttggaattata
 ggtatgatgaagaaaaaactaatactatagcaagtgggtgattttatattataggtgca
 15 atattaagtttaaatatgattacttttatatcttggttagtggtggaatattattaatt
 aaaaaagacaaaaataaaagcataaaagacaataaaacacatttggtggattag

Sequence 1536

MMNKAINITSLIGIILQSFSSLLFLVFLVFSITGAMDANFTTTVNGETTVHSAEAAQHVF
 20 SVGFLVIFIVAIFSIIFGIIGMMKKKTNTIASGVFYIIGAILSLNMITFISWLVCGILLI
 KKRQNKSIKDNKTHLVD*

Sequence 1537

Contig_0615_pos_3907_4716,
 25 is similar to (with p-value 1.0e-22)
 >gp:gp|U93876|BSU93876_19 Bacillus subtilis aminoglycoside 6
 -adenylyltransferase (aadK) gene, partial cds, and YrdA (yrd
 A), YrdB (yrdB), hypothetical protein YrdC (yrdC), YrdD (yrd
 D), hypothetical cytochrome P450 protein YrdE (yrdE), ribonu
 30 clease inhibitor (yrdF), regulatory protein YrdG (yrdG), hyp
 othetical protein YrdH (yrdH), hypothetical protein YrdI (yr
 dI), amino acid transporter (yrdJ), YrdK (yrdK), LysR family
 regulatory protein YrdL (yrdL), YrdN (yrdN), cation transpo
 rt protein YrdO (yrdO), hypothetical protein YrdP (yrdP), Ly
 35 sR family transcription regulator YrdQ (yrdQ), hypothetical
 protein YrdR (yrdR) and hypothetical protein YrkA (yrkA) gen
 es, complete cds. NID: g1934641. >gp:gp|Z99117|BSUB0014_140
 Bacillus subtilis complete genome (section 14 of 21): from 2
 599451 to 2812870. NID: g2634966.
 40 gtgacaatcttagcgattgatattggagtggaatgtgggaatagcatcagcaattgtaaca
 attgtgattatacttatttctgaagtgattcctaaatcaattgctgcaacatttctgat
 aaaatttcaaaacttggtatcctatcattcatatatgtgttattgtactcaagccatt
 acaatcttattaaacaagatgacagatggtattaatcatttactatctcgaggccaacct
 gttgaaaaaagattttctaaagaagaattcgtagcattattaaatattgctggtagagaa
 45 ggtgcatttaattgagatagaaaaatactcgacttcaaaacggttatggactttgaacaattg
 aagggttaaggatgttgataccacgcctcgtattaatgttgtagctttttcaaagggaagta
 acatatgacgaagcttatgatacagtgatgaataacccatatacaagatatccagtatat
 gatgaaaatatagatgatatcatcgccgtattccactcaaaatatttattagcttgaggt
 aaaaaataagaggacgcaattactaattatgcatcaagccctttatttgtaaatgaacat
 50 aatagggcagaatgggtattgctgtaaaatgaccgtttcacgaaaacatttagcgattggt
 ttagatgaatttgaggtacggatgctatcgatcgacgaagatttaatagaagagcta
 cttggtatggatattgaggatgaaatggatcgtgaagaagaaaataaattaaaacatcaa
 aaatttccgcaaagcatgatgcacgttaa

55 Sequence 1538

VTILAIDIGVNVGIAIAIVTIVIIILISEVIPKSIATFPDKISKLVYPIIHICVIVLKPI
 TILLNKMTDGINHLLSRGQPVKEKRFSEKIEIRTLNLIAGREGAFNEIENTRLQNVMDFEQL
 KVKDQDVTTPRINNVAFSKEVITYDEAYDTVMNNPYTRYPVYDENIDDIIGVFHSHYLLAWS
 KNKEDAITNYASSPLFVNEHNRAEWVLRKMTVSRKHLAIVLDEFGGTDIVSHEDLIEEL

LGMDIEDEMDREEENKLKHQKFPQSMHR*

Sequence 1539

Contig_0615_pos_5401_0,

5 putative peptide of unknown function

atggtgctcgtagacaatttccaccttggtttgattgtaacgtccaaaatataaattacatc
ttatatgtgagaaaacaattaactgatattccgatgagcattgaatttagacatcaatca
tggtttgacaatcagtataaagaacaaactttatccttcttaacacacatcaaatcatt
catgcagtggttagatgaacctcaagttaagaggggagcggttcctttagtaaataggatt
10 actagtgaattgcttttgtagcttatcatggacgtaatcattatggttggactaaaaaa
gatatgactgatcaagaatggcgagatgtaagatatattatgattatagcgatgatgag
ttagctgacttggtcgtgaaagtcgaaataacttaatcaaaaggctaagaaagta

Sequence 1540

15 MVLVQFPWFDCNVQININYILYVRKQLTDIPMSIEFRHQSWFDNQYKEQTLSTLQHQII
HAVVDEPQVKEGSVPLVNRITSEIAFVRYHGRNHYGWTKKDMTDQEWDRVRYLYDYSDD
LADLARKVEILNQKAKKV

Sequence 1541

20 Contig_0616_pos_8264_0,

putative peptide of unknown function

atgtccgcgtttattgaacaatctcaatatattgcgattcataatcaagataatttatat
gatgatttattccagtttttagtaaaaaataaaagatatctataaaacaaaactaggtagt
gctgtgattgaaatattaatttagtcatcaacaaatggaagctagagaaacttttatgact
25 aattactttaatcataatcgcaaagtttttaaagagattgttcgtaagcacatacaagag
gaagaacaagatttgtttattgatttaactcttcacccatctattttaatatattaatt
aaacctgaaactctggatga

Sequence 1542

30 MSAFIEQSQYIAIHNQDNLFDLQFLVKIKDIYKTKLGSAVIEILISHQQMEARETFMT
NYFNHNKRVLKEIVRKHIQEEEQDLFIDLIFSPIYFNILIKPETLDX

Sequence 1543

Contig_0616_pos_7255_6641,

35 putative peptide of unknown function

atgaatgctccattttattttaatagctgatcctagaatcgaagggtggtgccttttaccta
gggtcagagaattatgaacaggcaatccgtaagggtcatccaaaatgctttggattatttg
ggatttgcaacaaccaattaattctttctggattatcaatgggatcatttggcgcaactt
tattacgctacaaaattaaatccagcggctgttattgttaggaaaacctttgataaatctc
40 ggtactattgctaataatgaaactcggttcgccaacgattttggaacgtcacttgat
attttgcgattgaatcaaaatggcataactaacaagatgttggttcagttagataatcat
ttttggaagcaaattcagcatagtgatttgtcaatgaccacatttgcgattgcttacatg
gagcatgatgattatgacaaatatgcatttcaagatttattgcctgttcttacaaaacaa
catgcacgtgtgataagtaaaagaattcctggttagacataatgatgattctgctactgtt
45 actcattggtttattaatttttataatttaaatcatggaagagcgatttgggagggtaaca
catgcaagaagatag

Sequence 1544

50 MNAPFILIADPRIEGGAFYLGSENYEQAIRKVIQNALDYLGFANNQLILSGLSMGSFGAL
YYATKLNPAAVIVGKPLINLGTIANNMKLVRPNDFGTSLDILRLNQNGITNKDVVQLDNH
FWKQIQHSDLSMTTFAIAYMEHDDYDKYAFQDLLPVLTKQHARVISKRIPGRHNDDSATV
THWFINFYNLIMEERFGRVTHARR*

Sequence 1545

55 Contig_0616_pos_6597_6082,

putative peptide of unknown function

atgtatggtacaaaattacgtttttaatcaagataatatctattttgagaaccctttgatg
ccatccggtacaatcattcacagttggtatatgttaactgattttgcagaagaccgtgta
agccctaagctacctattttaaaaaaaggcgccaatatcaatttcaatttaattttgaa

gttgaacctgaggggtgcggttatttttaaaatgaaattttatcgtaagaataaagaaatt
 cttagtcatcaaattctaaaaataaaaaagaaaatattgtctatcctagagaagcatat
 tcatatgaattagaacttattaatgctggcatgaatcatctatcttttcacaataataatt
 gtgcaagaattaagagaagatagtaatcaagcttatgaggcaacgcaatatatagatcct
 5 aagaaaaaacttaaagtaattaatcaaataataaccaatataaggacacatcatctagac
 tcatcaaactatcacaggagtgatatgaatggctaa

Sequence 1546

MYGTLKRFNQDNIYFENPLMPSGTIIHSWYMLTDFaedrvSPKLPIlKKGRQYQFQFNFE
 10 VEPEGAAYFKMKFYRKNKEILSHQILKNKENIVYPREAYSYELELINAGMNHLSFHNII
 VQELREDSNQAYEATQYIDPKKKLKVINQIITNIRTHHLDSSNYHRSDMNG*

Sequence 1547

Contig_0616_pos_5858_3699,
 15 is similar to (with p-value 0.0e+00)
 >sp:sp|P47994|SECA_STACA PREPROTEIN TRANSLOCASE SECA SUBUNIT
 . >pir:pir|S47149|S47149 secA protein - Staphylococcus carno
 sus >gp:gp|X79725|SCSECA_2 S.carnosus (TM300) secA gene. NID
 : g499333.
 20 atgtatccaaagatgtgcagatttttaggagcaatcgctatgcacaggggaatattgca
 gaaatgcaaacaggagaaggtaagacgcttacagctaccatgcctctgtacttaaatgca
 cttacaggtaaaaggtgcttatctaatacacaacaaatgattacttagcaaacgcgatttt
 ttagaaatgaaaccactatatgaatggctaggcttgcctgtatcattaggatttggggac
 attccagaatatgaatacgcgtgaaaatgaaaaatatgaactgtaccaccatgacattggt
 25 tacacgactaatgggcgactagggtttgattatttaattgataatttagctgatgatatt
 cgtgccaaatttttaccgaaattaaactttgctattatttgatgaagtcgattctattata
 tttagacgctgccccaaacgccttttagttatttctgggtgcaccacgtgtacaatctaattta
 tttcacatcggttaaaaagtttgttgaaacacttgagaaagataaagacttcatagttaaat
 ttttaataaaaaagaagtgtggctcactgatgagggctcggaaaaagcaagccattatttc
 30 aaagtgaatagtatataccaacagcaatattttgatttagttaggatgattcatttatcg
 cttagagctaagtatttattcaaatataatttagactattttatttttgatggtgagatt
 gtgcttatagatagaataactggctcgatgctacctggaacaaagcttcagtcctgggtta
 catcaagctatagaggctctggaaaatgttgaaatttctcaagatatgagtgtgatgyca
 accataacattccaaaacttatttaagcaatttgatgaattttcaggtatgactggaaca
 35 ggtaaattaggggaaaaagaattctttgatatttatattcaaaagttgttatagagattccg
 actcacagtcggattgaacgagatgtagacctgatagagtatttgctaattggtgacaaa
 aagaacgatgcaattttaaagacagtgattggtatacatgaaactcaacaacctgtgta
 ctaattacacgtactgcagaagcggcagaatatttttcagctgagttatttaaacgtgat
 ataccaacaatttattaatcgctcaaaatgtagctaaagaggcacaatgattgctgag
 40 gcgggacaattattgcagttactgttgcataagatggcagggcggtggaactgatata
 aagttatcaaaagaggttcatgatatcggtggcttagcagtgattttaatgaacatatg
 gataatagccgtgttgatcgtaattaaaggagcgtcaggtcgccaaggagatcctgga
 tattcacagatttttgatcacttgatgatgatttagtaaaacgttggagtaactctaac
 ttggcagaaaaataaaaacctccaaacgatggatgcatctaaactagaaaagttagtcactc
 45 tttaaaaaacgtgtaaagtcaattgttaataaagcgcaacgtgtatctgaagagactgct
 atgaaaaatagagaaatggcaaatgaattcgaaaaaagtattagtgttcaacgagataaaa
 atttatgctgaacgtaatacacataacttgaaagcaagcgattttgatgatttttaattttgaa
 cagcttgacagagatgtgtttacaaaagacgttaaaaaatcttgacttaagtagtgaacgt
 gcacttgtgaattatatatacgaaaacttaagttttgtcttcgatgaagatgtatcaa
 50 attaatatgcaaaatgatgaagaaatcatacaattcttaatacaacaatttactcaacaa
 tttacaatcgtttgaagttgctgctgattcatatttaaaacttcgtttcattcaaaaa
 tcaattttgaaagcgatagatagcgaatggattgaacaagtagataacttacaacaactt
 aaagccagtgtaaaccaatcgacaaaatggacagcgtaatgtcatttttgaatatcataaa
 gtggtccttgaaacgtatgaatatatgtctgaagatataaaaaggaagatgggtagaat
 55 ttatgttttaagtattctagcctttgataaggacggagatatgggtcatttcccataa

Sequence 1548

MYPKDVQILGAIAMHQGNIAEMQTGEGKTLTATMPLYLNALTGKGAYLITNDYLAKRDF

LEMKPLYEWLGLSVSLGFVDIPEYEAENEKYELYHHDIVYTTNGRLGFDYLIDNLADDI
 RAKFLPKLNFIAIDEVDSIILDAAQTPLVISGAPRVQSNLFHIVKKFVETLEKDKDFIVN
 FNKKEVWLTDEGSEKASHYFKVNSIYQQQYFDLVRMIHLSLRAKYLKYNLDYFIFDGEI
 5 VLIDRITGRMLPGTKLQSGHLQAIEALENVEISQDMSVMATITFQNLFKQFDEFSGMTGT
 GKLGEKEFFDLYSKVVEIPTHSPIERDDRPDRVFANGDKKNDAILKTVIGIHETQQPV
 LITRTAEAAEYFSAELFKRDI PNLLIAQNVAKEAQMIAEAGQLSAVTVATSMAGRGTDI
 KLSKEVHDIGGLAVIINEHMDNSRVDRQLRGRSGRQGDGYSQIFVSLDDDLVKRWSNSN
 LAENKNLQTMASKLESSALFKRVKSIVNKAQRVSEETAMKNREMANEFKSIISVQRDK
 IYAERNHILEASDFDDNFQELARDVFTKDVKNLDSLSEALVNYIYENLSFVFEDEVSN
 10 INMQNDEEIIQFLIQQFTQQFNNRLEVAADSYLKLRFIQKSILKAIDSEWIEQVDNLQQL
 KASVNNRQNGQRNVIFEYHKVALEYEYMSEDIKRKMVRNLCLSILAFDKGDMVIHFP*

Sequence 1549

15 Contig_0616_pos_3690_2173,
 is similar to (with p-value 3.0e-20)
 >sp:sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) A
 LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TEICHOIC ACID BIOSYN
 THESIS PROTEIN E). >pir:pir|S06048|S06048 probable rodD prot
 20 ein - Bacillus subtilis >gp:gp|X15200|BSRODC_1 Bacillus subt
 ilis rodC operon. NID: g40098. >gp:gp|Z99122|BSUB0019_70 Bac
 illus subtilis complete genome (section 19 of 21): from 3597
 091 to 3809700. NID: g2636029.
 gtgttagacatgacgatttataatatcaattttggaatcggttggccagtagtggtgtt
 25 gaatatgcacaagtgtatcgagcaaaactattaaggcaattaccttattccaacaaaattt
 atatttttagattttattcaatcagaaaaatattcaaacactcacaagcaacatagggtt
 aaagatgatgaagtattttggctatatcaatacttcacagacgtaaaaaatcgctcctaca
 acgtacacagttgatgtatttaatttcagagtttaggtaatgaggttactcgaaaagaacaa
 aacggtaaagtattacgactctatttaataatcagcaaacattcgtaacctgctattta
 30 aaaaatgctaacgaacattacgttgatcggtgcagagtttggtggaatggaatgttaatt
 aggaaagatttttatagctatgtaagaacattctcgaggtattatgctccttttaacaat
 aaagctaaaaatataatgctcaattttataatgaaaatggatcaattgcatatcgcgaa
 tatatagatgaagatgaacatgtttttgtgtttgatgatgcagtttatcagtaaacaa
 gcaactcgctgcatactttatacgtcaatttgctattgaattcagaggatattattatt
 35 gatcgagcaactgatgtgggacaagccatattagaaaaaaggatccagtaaaagttggc
 gtatgctggttcacgctgaacatttcagtgaaggtgagactgatgggactcacattttatgg
 aataattattatgagatcaattcgaaaatgcacagcatattgatttctttatcacagcg
 acagattttgcaaaggcagacactaagtgaacaatttaaacaaatataagaatgattgtcca
 cgtatacgtacaattccagtaggtagtagatagaatcgttacaatatcctgaaaaagaaaga
 40 aaaccatatttccatcatgaccgcatcacgtcttgctaaccgaaaaacatgttgattggata
 gtggaagctgtgatttaaaagctaaacatcagttacctcaattgagttttgatattacgga
 caaggagaagaacaagaaaaattaaaaatattattaccaaacatcggtgctgaggattac
 atacaaattaaaggacatagaaatcttcgtacaatatatcagcaatatgaattattcata
 gcggcctcacaaagtgaaggggttcggactgacattaatggaagcggttggtctgtggtta
 45 ggtatgattggatttgatgtgaattatggcagtcggacatttattcgacatcatcaaat
 ggctatttgataccgatagattttgaacaagcgtctactgatgatcacaaacgcaaatc
 gctcatatgattattcgatattttgaagatggtcccataaaggccacatgaggtatcgat
 gacattgcagaatcatttaaacatcgcatattgttgatctatggagacaactcattgaa
 gaggtgctatatgattaa

Sequence 1550

50 VLDMTIYNINFGIGWASSGVEYAQVYRAKLLRQLPYPTKFIFLDFIQSENIQTLSNIGF
 KDDEVIWLYQYFTDVKIAPTYYTVDDLISELGNEVTRKEQNGKVLRLYLNNQTFVTCYL
 KNANEHYVDRAEFVNGMLIRKDFYSYVRTFSEYYAPFNNKAKIYMRQFYNENGSIAYRE
 55 YDEDEHVVFVDDARLYSKQALVAYFIRQLLLNSEDIIIDRATDVGQAILENKGSSKVG
 VVVHAEHFSEGATDGTHILWNNYYEYQFENAQHIDFFITATDLQRQTLSEQFKQYKNDGP
 RIRTI PVGSIESLQYPEKERKPYSIMTASRLANEKHVDWIVEAVIKAKHQLPQLSFDIYG
 QGEEQEKIKNIITKHRAEDYIQIKGHRNLRTIYQQYELFIAASQSEGFGLTMEAVGSGL
 GMIGFDVNYGSPTFIRHHQNGYLIPIDFEQASTDDITQIAHMIIRYFEDGPIRAHEVS

DIAESFKTSHIVDLWRQLIEEVLYD*

Sequence 1551

Contig_0616_pos_1748_840,

5 putative peptide of unknown function

atgacacgtgaaggtaaagaagtgatttatgagaattatgtaactaacgatgtagttgta
gaatatgaagggaaatcttatttttttgagtcataacagagtggttaaattttacttg
agtgaatgggcattgagataaaagaagttatatttaatactttatcaacaccattttta
gcaatttatcatttgccgacattgaaaaaagggtatttttttggcaagaacaatctcag
10 ggttatgtcccaggaaatatgaaagtcattgttatcaccaaacccttcaaagtcgctttgcc
gttattgtccctaatacagaatgaatacaaatgatcaaggaaacaactatctaggaggaa
caacaggcagcatatgcatctggttacttatatgacacgtataaacggaatcattattct
aagaatgtattaacattaacaaattcagatcaaataccacatgttgaaacgttggtacgt
ttgcataaagattatcaatttcacataggcgctaaaactgagatgtcttcaaaattatta
15 agtttatcgcaatatgaaaatgttaaattatatccaataattaaagaacaaacagttcaa
accttatatcaacaatgtgacatctatttagatattaatgaggggaacgaaatagggaat
gctgtaagaagcgcatataatcatcaattgttaattatgggatataaagaggttggtcat
aatcaagatttcggtgcaatgaaaaatcagtttcttgtaaatgataaagtcagttgagt
aacgctttgaaagagataggaaatcatcggtggtcaatttgaaacagtttagcactacaa
20 caacgtcatgctaattgctgtgccggtatcaacatttaatacgcattagtacaagcatta
agtgggtaa

Sequence 1552

MTREGKEVIYENYVTNDVVVEYEGKSYFFESYTEWIKFYLSEMGIEIKEVIFNTLSTPFL
25 AIYHLPFLKKGILFWQEQSQGYVPGNMKVMLSPNLQSRFAVIVPNQNEYKLIKEQLSREE
QQAAYASGYLDYTYKRNHYSKNVLTLTNSDQIPHVETLVRLHKDYQFHIGAKTEMSSKLL
SLSQYENVKLYPIIKEQTVQTLYQQCDIYLDINEGNEIGNAVRSAYNHQLLIMGYKEVVH
NQDFVAIENQFLVNDISQLSNALKEIGNHRGQFETRLALQQRHANAVPVSTFKYALVQAL
SG*

30

Sequence 1553

Contig_0616_pos_734_387,

putative peptide of unknown function

atggaaaattttgataaaagtattcatgataaaacgggtgatgtattaggggctttaagt
35 tatctaagtgtatttttcgcacctgtattgtttccattaatcgtatggattgttgacaa
ccaccagcatctacgtattcaagaaatgcattatttaaccatattttgagttgggtgtgt
ttggtattaggacttatatcatttgctgctggactatccttgattgattcgacaaatgga
gtcgctgtactagtgtataggagtaattattggaggtattctacttatcgcttcgcttgta
40 ttattattattaatatgtgaagggtatcaaattattgatgatatag

Sequence 1554

MENFDKSYHDKTGDVLGALSYSVFFAPVLEPLIVWIVGQPPASTYSRNALFNHILSWVC.
LVLGLISFAAGLSLIDSTNGVAVLVIGVIGILLIASLVLFIIINIVKGIKLLMI*

45 Sequence 1555

Contig_0618_pos_2737_2393,

putative peptide of unknown function

gtgacaaaccggaggaaggtggggatgacgtcaaatacatcatgcccttatgatttgggc
tacacacgtgctacaatggacaatacaagggcagcgaaccgcgaggtcaagcaaatcc
50 cataaagttgttctcagttcggtattgtagtctgcaactcgactatatgaagctggaatcg
ctagtaatcgtagatcagcatgctacggtgaatacgttcccgggtcttgtacacaccgcc
cgtcacaccacgagagtttgtaacaccgaagccggtggagtaaccatttggagctagcc
gtcgaagggtgggacaaatgattgggggtgaagtcgtaacaaggtag

55 Sequence 1556

VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSSDCSLQLDYMKLES
LVIVDQHATVNTFPLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVTR*

Sequence 1557

Contig_0619_pos_1598_3097,

is similar to (with p-value 0.0e+00)

>gp:gp|D50097|D50097_1 Bacillus subtilis macronuclear gene f
or lactate permease, complete cds. NID: g2258092.

5 atgaaaggaatttatgctgctatcacacccttggtgttacattattaattgcaattcca
ttctttaaattaccagtaggaattgcctctggagcagttggtgaagggtttctccaaggt
atcttcccaatcgatataattgttattatggcagttattatataagattactttgaaa
tcggggcaattcgcaactattcaagacagttattacaagttttcacagaccaaagaatt
cagcttcttttaattggtttttcatttaaatgcattcttagaaggcgctgcaggatttggg
10 gttccaattgcaatttgtgcaacttttattagcgcaacttggcttttagaccattacaagca
gctatgttatgttttagtagctaacgctgcattctggtgcatttgggtgcaattgggtattccg
gttgggtgttagatcacacttaacttacctggctatgtagaagcgatgggagtttcacaa
acatcaacattaaacttttagcaattattaacttctttattcctttcttacttatctttatc
gtagatggtttcaaaggaattaaagaaactttaccttcaattcttgttgtttctgtcact
15 tatacagttttacaaggattacttacagtgtttaatgggtccagaattagctgatattcatt
ccatcacttgccttctatgttagcatttagctttattctctaagaaattccaacctaagaat
atcttttagagttcaaaaagatgttaaaccagaagcacggaaaaaacttaaggtaaagaa
gtcttatttgccttggagtcattcattatcttaactgtcattgttatgatttggagtgca
ccttcattttaagcattatttgcaccaaaggtaaatctctgctttagttgcaaaacttt
20 gacttacctgggtactttcagtaaatattcacacaaaccaattactttatcattaaactta
attgggtcaaacaggtacagcaattctaattacaattatttactgttttaattggctaaa
aaagtcaactttggtgatgctggcgcttatttgtgaagcatttaaagaattatgggta
ccaatcataacaatttgtttcatcttagcaatttcaaaaatcacacatatggtgggtta
agtaatgctatgggacaaggcatctcaaaagcaggaagcgatttcccaatattatcacca
25 atccttgggtggatcggtctatttatgactggttcagttgttaataacaactctttattc
gcgccaattcaagcttctgtagcacacaaattggtacaagtggttcactactttagct
tcaaatacagcaggtggggttgcggcgaaacttatttctccacaattctattgccattgca
acagcagctgttaaagaagtaggtaagaatctgaactacttaaaatgacattacgttat
agtattggattacttgtatttatctgtatctggacatttatcttgcattcattctgtaa
30

Sequence 1558

MKGIYAAITTLVVTLIAIPFFKLPVGIASGAVVEGFFQGIFFPIGYIVIMAVLLYKITLK
SGQFATIQDSITSISQDQRIQLLLIGFSFNAFLEGAAGFGVPIAICALLAQLGFRPLQA
35 AMLCLVANAASGAFAIGIPVGVVDTLNLPGHVEAMGVSTSTLTIAIINFFIPFLIF
VDGFKGIKETLPSILVSVTYTVLQGLLTVFNGLPELADIIPSLASMLALFSSKKFQPKN
IFRVQKDVKEAPKKLKGKEVLFWSPIILTVIMIWSAPSFKALFAPKGLSALVANF
DLPGTFSNISHKPITLSLNLIGQTGTAILITIIITVLMAKKVNFGDAGRLFVEAFKELWL
PIITICFILAIKITYGGLSNAMGQISKAGSVFPIISPILGWIGVFMGTGSVVNNNSLF
40 APIQASVAQQIGTSGSLLVASNTAGGVAAKLISQSIATAAAVKEVGKESELLKMTLRY
SIGLLVFICIWTFILSFIL*

Sequence 1559

Contig_0619_pos_6750_6403,

45 putative peptide of unknown function

atgtttactgaattagaagatttgtatttccgaggaaatgaaagactattacatcaagct
ttcaataacctcatcattaatgcaattatgctcctcaaaatagcatgattaatatac
actctaactagtacaaatcacttgattatatttaattgaaaatgatggatcgattgca
gaagaagatgcgaaacatatcttcgatcggtttttataaactgagtgaacgaatctagtagt
50 aatgggtctagggtctagccattacccaatcaatcattcatcttcatcatggtagcattact
ctcacttcagatgataaaacacaatttattgtaaaactattttatttag

Sequence 1560

55 MFELEDLYFRGNERLLHQAFNNLIINAMNYAPQNSMINITLTSTNHLIIFNIENDGSIA
EEDAKHIFDRFYKLSDESSNGLGLAITQSIIHLHGSITLTSDDKTQFIVKLF*

Sequence 1561

Contig_0619_pos_4428_3304,

is similar to (with p-value 5.0e-37)

>gp:gp|AL023702|SC1C3_12 Streptomyces coelicolor cosmid 1C3.
NID: g3169026.

atgcataatattttatgcaatgggcggaacggtaaaagtcggtgacacaacttgcaaataca
ctggcagaaaaaggacatcctgtaacaattatttcagtttttagaggcgagactctcca
5 tattttgaattacattcagcaataaaagttaaagtcgtagtgactatcgcttaaagctt
aaaaatactagagctattacggcaaatcgatcaaaaagtataccccctttttaataca
aaagtgattttctcaatttgagccaggtaaaagtcagttttcgagttatgtagagaaaaaa
atgattaaagcaatcaggcatactaaaactgatgtactcggtggaacaagagctagcttt
aatattctcattttcaaatatgctaaagctgaaatagtgaccatcgcaatggaacatatg
10 aattttgatgctcaccctgatcagtatcaaaaggaaattattgctgctgacccgtaatatc
aataagattacaacgttaactgtcgcagaccagcaaaaatatcaatcacaaacttaaaact
cctgtatacgttatacctaataatggttaccgaaaaaagaattgctgctccaaagaaaaat
cgtattattagcgcgggacgtttagaatatgaaaaagggatgatttattattagagagt
attcgtttaatacaagaagacttgcgtcaattgaattatgacgttcacatctatggttct
15 ggtagtaagaaaaacatcacttgttgactttattaatcaatatcatttaaagatttgatt
aaaatatatgagccaacgcaagaattaaataataaacttgacaaaagtaaaatcggtgtt
gtaccttcacgcaatgaaggtttcggaatgattatttagaggcaatggtgcaagataat
atagtaataagttttgaaggcaatgtagggccagattcaatcattaacaacggagataat
ggttatttagtaaacctatgaaaatgtgtctgaacttgcaaacgtatcgatttaacaaca
20 caacattataatgagttagatcacatcattgaaaatagtaaagatcggtgaaacaattt
agtccggatcatatatatcaattatttatgtctatgtttaataaa

Sequence 1562

MHNIYAMGGTVKSVTQLANTLAEKGHPVTIISVFRGADSPYFELHSAIKVKVVVDYRLKL
25 KNTRAITANRIKKYTFELNTKVISQFEPGKSQFSSYVEKKMIKAIRHTKTDVLVGTRASF
NILISKYAKAEIVTIAMEHMNFDAHPDQYQKEIIAAYRNINKITTLTVADQQKYQSQLKT
PVYVIPNMVTEKRIAAPKKNRIISAGRLEYEKGYDLLESIRLIQEDLRQLNYDVHIYGS
GSKKTSLVDFINQYHLNDLIKIYEPTQELNNKLAQSKIVVPSRNEFGMIILEAMVQDN
IVISFEGNVGPDSIINNGDNGYLVNYENVSELAKRIDLTQHYNELDHI IENSKDTLKQF
30 SPDHIYQLFMSMFK*

Sequence 1563

Contig_0619_pos_0_1148,
putative peptide of unknown function

atgataaagcaataaataattttcaaatatggacaaattaaaagagcaaatggaacgtgca
35 cttagcgacggctatacgcattgtcatcccctattcaaatgaaattcaaatcatcagtc
atgattaaagctatcactttacctaagacttcatttatagttgactatacaattaacaat
tattattttaaacgattgtaataacttcgggttgactttgttgattttgaggactgggtt
aaaaatatttaattatatccaaatgttatttatgaaattaattcaacattagaacttatt
40 gataaatttgaaagttgaaaatatctttgatttagcattattaacaattcttaaaggcat
atcgcagttgaaggctatgtcgtatttagactttaaaggaccattaaaaacgagcaaggga
ttttggcgctcatttgaccgtaatgatttaacttatagagataaattcttcttaaacacc
atcgcttatgcacataaacaagaatcccatttacgcgtgtaccatttaacgatcacgat
agtattagatattatgattcagttactacttagtactaaatttaaagctccaagatggtta
45 gtgactcctattaagaattattcagttaaaaaacacaaagagattagctatatttataaa
aaggattcatcaaaaacttaagaaccacgtcggttttctaggatttgatttcggctatcga
ggaaactctaagttatttatttaattactttgttaaacacaatcctatgatagagtccttac
ttataacagatgagagaacaggaccacattttatttcaactaacgatgaaaatgtgaag
aatttgattgaaacagctactttgtcattacggaaagctatattcctgacgacattcac
50 cctaattggtaaaatcatccaattatggcatgggacacctattaaaaaactatttttagat
agtaaagagccacacaaaatttaaatatataaactaccgagctcgaaaatataataaa
tggaacacaagaattatttaattgtagattcagaagaatcaaaaacatactttgaatca
gcgtttcctagtcaaaaaattgatataattacctgtaggatatcctagaataaattattta
55 ttaaatAA

Sequence 1564

MIKQINISNMDKLKEQMERALS DG YTHVIPYSNEIQIHQSMIKAITLPKTSFIVDYTINN
YYLNDCKYFGLDFVDFEDWVKNINLYPNVIYEINSTLELIDKFEVENIFDLALLTILKGH
IAVEGHVVLDFKGLKTSKGFWRSFDRNDLT YRDKFFLNTIAYAHKQRI PFTRVPFNDHD

SIRYYDSVLLSTKFKAPRWLVTPIKNYSVKKHKEISYIYKKDSSKLKNHVFLGFDGGR
GNSKYLNFYFVKHNPIMESYFITDERTGPHFISTNDENVKNLIETATFVITESYIPDDIH
PNGKIIQLWHGTPIKKLFDSKEPHQNLNIYNYRARKYNKWTQQDYLVLDSEESKTYFES
AFPSQKIDILPVGYPNNYLLNX

5

Sequence 1565

Contig_0620_pos_622_942,

putative peptide of unknown function

atgactaaaattagtggtgtcgatatatggagcagaagtcggttggtgaggttggtgtaaat
10 gcacctacatctatagatacttatcaatggcttcaagcattacttttaagaaagtttcct
caacatcattttgaatttacatataattgacatacgaatgatactgaaaatttaactgat
catgatatgcaatttatagaagaattaatgaagatgaattgttttaccattagttacg
atgaatgatgaatatgtagcagatggttacatacaataataaacaataaaccgttttatt
aaatcatattttactatgtaa

15

Sequence 1566

MTKISVVVYGAEVVCASCVNAPTSIDTYQWLQALLLRKFPQHHEFTYIDIRNDTENLTD
HDMQFIERINEDELFPVLTMMNDEYVADGYIQYKQITRFIKSYFTM*

20

Sequence 1567

Contig_0622_pos_6645_7433,

is similar to (with p-value 4.0e-58)

>sp:sp|P45024|YA80_HAEIN_HYPOTHETICAL_AMINO-ACID_ABC_TRANSPOR-
25 TER_BINDING_PROTEIN_HI1080_PRECURSOR. >pir:pir|I64181|I6418
1 glutamine-binding periplasmic protein (glnH) homolog - Hae
mophilus influenzae (strain Rd KW20) >gp:gp|U32788|U32788_5
Haemophilus influenzae Rd section 103 of 163 of the complete
genome. NID: gl574629.

atgaaaagacttttacttttgattgttgacttgtttttgttttagcagcctgtggcaac
30 aattcatctaacaataaagataatcaatcaagcagtaaaagacaaggatacgttaagagtt
ggtacggaaggtacatatgcgccctttacttaccataataaaaaagatcaattaacaggt
tatgatattgatgtgattaaagcagttgcaaaagaagaaaatcttaacttaagttaaat
gaaacgtcatgggattcaatgtttgcaggattagatgctggtcggttttgatgttattgca
aatcaagtggtgtgaataaagatagagagaaaaaatataaattctctgaaccttacaca
35 tattcaagtgtgtacttgttgttcgtgaaaaatgaaaaagataattacatcattcaatgat
gtaaaaggtaaaaagtttagcacaacgtttacgtctaattatggtcaattggctaaagat
aagggtgctggagcgttactaaggttagatggatttaataatcaatcaatggacttactattatct
aaacgtgtagatggttacatttaacgacagtttatcttacttagattacagaaaaacaaag
cctaagtctaaaattaaagcaatcaaaaggacatgcagagcaaaataaatcagcatttgca
40 ttctctaagaaggttgatgaaaaacgattgagaaatttaataaaggcctagaaaaaatt
agagataatggtgaatttagctaaaattggttaagaaatggtttggtcaagatgtttctaaa
cctgaataa

Sequence 1568

45 MKRLLLCIVALVFVLAACGNSSNNKDNQSSSKDKDTRLRVGTEGTYAPFTYHNKKDQLTG
YDIDVIKAVAKEENLKLKFNETSWSMFAGLDAGRFDVIANQVGVNKDREKKYKFSEPYT
YSSAVLVVRENEKDITSFNDVKGKKLAQTFTSNYQQLAKDKGADVTKVDGFNQSMOLLLS
KRVDGTFNDSLSYLDYRKQKPNKAIKAIKGHAEQNKSAFAFSKKVDEKTIEKFNKGLEKI
RDNGELAKIGKKWFGQDVSKPE*

50

Sequence 1569

Contig_0622_pos_7495_8133,

is similar to (with p-value 6.0e-57)

>sp:sp|P45023|YA79_HAEIN_HYPOTHETICAL_AMINO-ACID_ABC_TRANSPOR-
55 TER_PERMEASE_PROTEIN_HI0179. >gp:gp|U32788|U32788_4 Haemoph
ilus influenzae Rd section 103 of 163 of the complete genome
. NID: gl574629.

gtgaaatactcaatccctattacttttggtcactttcattctaggtttaatcattgcattg
tttactgcacttatgcgtatatcaaccagtaaatgtcttagaggtattgcgcgtgtctat

gtatcaattattcgtggtacacctatgattgtacagttatttattattttttacggtata
ccggagcttggagattggttaactaacaatgctgataatcaatggacacttgcacctgtt
attgctgcagtcattggtttatctttaaatgttggtgcttatgcttcigaaattatacga
ggtggtatattgtctattcctaaaggacaaacagaagcggttactctataggtatgaac
5 tatagacaaaactgtgcaacgtattatcttaccacaagctattcgtgtatctataccagca
ctaggaaacacattttttaagtttaattaaagatacatcattacttggattttattcttgtt
gcagagatgtttagaaaggcacaagaagttgcttcgacaacgtatgagtatctaactatt
tatttggtagtagctttaatgtattgggtcgtatgttttgtcatctcaattatccaagga
tggtatgaatcacgcattgaaagagggtatcgctcatga

10

Sequence 1570

VKYSIPITLVTFILGLIIALFTALMRISTSKLLRGIARVYVSIIRGTPMIVQLFIIFYGI
PELGRLVNADNQWTLAPVIAAVIGLSLVNGAYASEIIRGGILSIPKGQTEAAYSIGMN
YRQTVQRIILPQAIRVSIPALGNTFLSLIKDTSLLGFILVAEMFRKAQEVASTTYEYLT
15 YLLVALMYWVVCFVISIIQGWYESRIERGYRS*

Sequence 1571

Contig_0622_pos_8265_8861,
is similar to (with p-value 6.0e-59)
20 >sp:sp|P39456|YCKI_BACSU PROBABLE AMINO-ACID ABC TRANSPORTER
ATP-BINDING PROTEIN. >pir:pir|S52383|S52383 probable ATP bi
nding protein - Bacillus subtilis >gp:gp|X77636|BSPAAT_3 B.s
ubtilis putative amino acid transporter gene. NID: g666980.
atgattaatgctttagagatacctactgaaggtacagtgtatgtcaatggcatgacatat
25 aatgctaagataaagaattctcaaattaaagtaagacaacaatcaggaatgggttttcaa
aattataattttatttccacataaatctgcattagaaaacggttatggaaggctctataaca
gttaaaaagatgaataaagcaacggctaataagaagcaatgaattttattgtctaagggt
ggattggtacatgttaagatcaacggccacatgctttatcaggagggaacaacaacgt
gtcgcaattgcacgtgcattagccatgaatcctaaagtgtattttgatgagccaaca
30 tctgcgcttgatcctgaattggtcaatgatgtattaaaagtcataaagaattggctgac
gaaggtatgacaatggtcattgtgactcacgagatgcgttttgccaaagaagtttccaat
caaattgcttttattcatgagggtgttattgcagaacaaggtagcctgaagatatattt
aatcatcccaaaacagaagagcttcagcgatttttaaatgtgattaatgaaaaatag

Sequence 1572

MINALEIPTEGTVYVNGMTYNAKDKKSQIKVRQOSGMVFQNYNLFPHKSALENVMEGLIT
VKKMNKATANEAMNLLSKVGLVHVVDQRPALSGGQQQRVAIARALAMNPKVMLFDEPT
SALDPELVNDVLKVIKELADEGMTMVIVTHEMRFAKEVSNQIAFIHEGVIAEQGTPEDIF
35 NHPKTEELQRFLNVINEK*

40

Sequence 1573

Contig_0622_pos_4380_4018,
is similar to (with p-value 3.0e-43)
45 >sp:sp|P46899|RL18_BACSU 50S RIBOSOMAL PROTEIN L18. >gp:gp|L
47971|BACRPLP_10 Bacillus subtilis ribosomal protein (rplPNX
EFROQ, rpmCDJ, rpsQNHEMK) genes, integral membrane protein (secY)
gene, adenylate kinase (adk) gene, methionine aminopeptidase (map)
gene, initiation factor 1 (infA) gene, RNA polymerase alpha (rpoA) gene.
NID: g1044970.
50 atgatcagtaaaattgataaaaacaaagtcagtttgaaaagacatgctcgtgttctgact
aaattatcaggtacagctgaaaagccacgtttaaatgtgtatcggttcaaacaacacatc
tatgcacaaattattgatgacgttaaaaggcgttaacacttgctcaagcatcatcacaagat
aaagatatgtgcaaacacatcagcttcaaaagttgacttagcaactactgttggtcaagaa
attgctaaaaaagctaacgataaaggtattaaagaaatcgtcttcgatcgcgaggatat
55 ttataccacggacgtgttaaagcttttagctgatgctgcaagagaaaatggattagaattt
taa

Sequence 1574

MISKIDKNKVRLKRHARVRTKLSGTAEKPRLNVYRSNKHIYAQIIDDVKGVTLAQASSQD

KDIANTSASKVDLATTVGQEIAKKANDKGIKEIVFDRGGYLYHGRVKALADAARENGLEF
*

Sequence 1575

5 Contig_0622_pos_3998_3498,
is similar to (with p-value 3.0e-53)
>sp:sp|P21467|RS5_BACSU 30S RIBOSOMAL PROTEIN S5 (BS5). >gp:
gp|L47971|BACRPLP_11 Bacillus subtilis ribosomal protein (rp
10 lPNXFROQ, rpmCDJ, rpsQNHEMK) genes, integral membrane prote
in (secY) gene, adenylate kinase (adk) gene, methionine amin
opeptidase (map) gene, initiation factor 1 (infA) gene, RNA
polymerase alpha (rpoA) gene. NID: g1044970. >gp:gp|Z99104|
BSUB0001_133 Bacillus subtilis complete genome (section 1 of
21): from 1 to 213080. NID: g2632267.
15 atggctcgtagagaagaagaactaaagaatttgaagaacgcgttggtacgattaaccgt
gttgctaaagttgtaaaaggtggtcgtcgtttccgtttcactgcattagtggtgttgga
gataaaaatggtcgtgtaggtttcgggtactggtaaagcgcaagaggtaccagaagctatc
aaaaaagctgttgaaagcagctaaaaaagatttagtagttgttccacgtgtagaaggtacg
actcctcactataactggtcaatatgggtcaggtagcgtatttatgaaaccagctgca
20 cctggtacaggagttatcgctggtggaccagttcgtgccgtattagagttagcaggaatt
actgatatcttaagtaaatcttttaggatcaaataatcctattaatatggttcgtgcgact
atcaacggtttacaaaacttaaaaaatgcagaagatggttgctaaattacgtggcaaatct
gtagaagaattatacaattaa

25 Sequence 1576

MARREEETKEFEERVVTINRVAKVVKGRRFRFTALVVVGDKNGRVGFGTGKAQEVPEAI
KKAVEAAKKDLVVVPRVEGTTPTTITGQYGSVFMKPAAPGTGVIAGGPVRAVLELAGI
TDILSKSLGSNNPINMVRATINGLQNLKNAEDVAKLRGKSVEELYN*

30 Sequence 1577

Contig_0622_pos_2773_1550,
is similar to (with p-value 0.0e+00)
>sp:sp|Q05217|SECY_STACA PREPROTEIN TRANSLOCASE SECY SUBUNIT
. >pir:pir|S30115|S30115 secY protein - Staphylococcus carno
35 sus >gp:gp|X70086|SCSECY_1 S.carnosus secY gene. NID: g49188
atgtagttatttttaaaataggaacgtatattcctgctccaggagttaatcctgaagcc
tttaatcatccacagggatctcaaggtgccactgagttattaaatacttttgggtggcggg
gccttgaaacgtttctcaatatgtgcgatgggaatcatgccttatatcactgcatccatc
40 gtcattgcaattactgcaaatggatattgttcctaaatttacagagtgggcaaaacaaggt
gaaatgggtagaagaaaaatttaataacgtaactcggtatttttgctataatttttagctttt
atccaatctataggtatggctttccaatttaataactatctcaaaggacaacttattata
gaaaagtcgtttatgagttatttattaattgcagttgtattaacagcggaacagctttc
ttaatttggttggtgaccaaatacacacagtttggtggtgtaacgggtatttctcttatc
45 atctttgcaggtatattatcaactttaccttcgagtcctagaacaatttgcaaatcagtg
tttgtaggtcaagacgatacttcaacttgcttggtgaaaatactaggattgattgtagcc
ttgattttactaacagtaggcgcaatatattgttcttgagagtaaacgtaaaatacctatt
caatatgcaaagaacaactctgctcaacgattaggttcacaagcaacttatctaccttg
aaagttaactctgccgggtgttattccagttatctttgcgatggcgtttttctgttacca
50 agaacattgactttattcttcccgaagcagaatgggcacagaatattgctgatactgcc
aaccatcaagtaatatgggaatgattttatgtagttttaattattgcatttgcatat
ttttatgcttttgtaacaagttaatcctgaaaaaatggcagataaccttaaaaagcaaggt
agttatgtcccaggaattagacctggtgaacaaacaaaaaatatattactaaagtactt
tatagattgacttttgggttcaattttcttagcagctatagctattttacctataatt
55 gcgactaaatttatgggttaccacaataatcgaattggtggtacgagtcctttgatc
gttattggtgtagctattgaaactatgaaaactttagaagcacaagtcactcaaaaagaa
tataaaggctttggtgtagataa

Sequence 1578

MLVIFKIGTYIPAPGVNPEAFNHPQGSQGATELLNTFGGGALKRFSIFAMGIMPYITASI
 VMQLLQMDIVPKFTEWAKQGEGRKINNVTTRYFAIILAFIQSIGMAFQFNLYLKGQLII
 EKSVMYSYLLIAVVLTAGTAFLIWLGDQITQFGVNGISLIIFAGILSTLPSSLEQFAQSV
 FVGQDDTSLAWLKILGLIVALILLTVGAIFVLEAKRKIPQYAKKQSAQRLGSQATYLP
 5 KVNAGVPIPVIFAMAFFLLPRTLTLFFPKAEWAQNIADTANPSSNIGMIIYVVLIIAFAY
 FYAFVQVNPEKMADNLKKQGSYVPGIRPGEQTKKYITKVLRLTFVGSIFLAAIAILPII
 ATKFMGLPQSIQIGGTSLLIVIGVAIETMKTLEAQVTQKEYKGFGR*

Sequence 1579

10 Contig_0622_pos_1532_885,
 is similar to (with p-value 2.0e-82)
 >sp:P16304|KAD_BACSU ADENYLATE KINASE (EC 2.7.4.3) (ATP-A
 MP TRANSPHOSPHORYLASE). >pir:pir|JS0492|JS0492 adenylate kin
 ase (EC 2.7.4.3) - Bacillus subtilis >gp:gp|L47971|BACRPLP_1
 15 5 Bacillus subtilis ribosomal protein (rplPNXEFROQ, rpmCDJ,
 rpsQNHEMK) genes, integral membrane protein (secY) gene, ade
 nylate kinase (adk) gene, methionine aminopeptidase (map) ge
 ne, initiation factor 1 (infA) gene, RNA polymerase alpha (
 rpoA) gene. NID: g1044970. >gp:gp|Z99104|BSUB0001_137 Bacill
 20 us subtilis complete genome (section 1 of 21): from 1 to 213
 080. NID: g2632267. >gp:gp|D00619|BACSECY_4 Bacillus subtili
 s genes for ribosomal proteins, SecY, adenylate kinase and m
 ethionine amino peptidase, complete cds. NID: g216336.
 atgaatatcatTTtaattgggcttacctggtgcaggtaaagggactcaggcgagtgaaatt
 25 gttaagaaattcccaataccacatatcttactggtgacatgttcagaaaagcgattaa
 gatgaacagatttaggaaaagaagctaaatcatatatggatcgtggagaattagttcct
 gatgaagtactgtaggtatcggttaaagaaagaatttctgaagacgatgcaaaaaaagga
 ttcttgttagatggattcccaagaactatagatcaagctgagtcattaagtcaaattatg
 tctgagcttgatagagaaattgatgctgtcattaatatcgagttcctgaggaagaatta
 30 atgaatcgtcttacaggtcgtcgatatctgtgagaaatgtggtacaacatatcatcttgta
 tttaatcctccaaagggttgatggtatatgtgatatcgatggtggaaagttatatcaacgt
 gaagatgacaatccagaaacagtatctaactggttgagcgtaattgtaaacatctaag
 cctattttagaattattacaacaacaagggtgtcttgaaaaacattgatggttcaaaagat
 attgacgaagtaaccaacgatgtcattgatctcttagatcatttataa

Sequence 1580

MNIILMGLPGAGKGTQASEIVKKFPIPHISTGDMFRKAIKDETDLGKEAKSYMDRGELVP
 DEVTVGIVKERISEDDAKKGFLLDGFPRTIDQAESLSQIMSELDREIDAVINIEVPEEEL
 MNRLTGRRICEKCGTTYHLVFNPPKVDGICDIDGGKLYQREDDNPETVSNRLSVNVKQSK
 40 PILEYYNNKGVLKNIDGSKDIDEVTNDVIDILDHL*

Sequence 1581

Contig_0623_pos_9597_9896,
 putative peptide of unknown function
 45 gtggtggtactgctgctgaacttgtgcccagaaaatcaacgcctaaatattgtcttttat
 ctgcccacatttgacgaacttccattgtagagggacttgatttgactttctaaagcaa
 caatatcgacattcatatcactatctaatacaacttttacagcctcaggatcccaaatg
 cattccattcagctgtaccatcatgttctggttcttcaacatttcctttgtctaaaaacg
 50 ttccacccatccaaactaatttctctatatattttttaaaattgagttgtcatattttatag

Sequence 1582

VVLLRLNLCPENQRLNIVFYLPFTDELPLLEGLDLYFLKQQYRHSYHYLIQLLPQDPKM
 HSIQLYHHVLVLQHFLCLKTFHPSKLISLYFLKLSCHIL*

Sequence 1583

Contig_0623_pos_13964_14308,
 putative peptide of unknown function
 gtgacaaaccggagggaagggtggggatgacgtcaaatacatcatgcccttatgatttgggc

tacacacgtgctacaatggacaatacaaagggcagcgaaaccgcgaggtcaagcaaattcc
cataaagttgttctcagttcggattgttagtctgcaactcgactatatgaagctggaatcg
ctagtaatcgtagatcagcatgctacgggtgaatacgttcccgggtcttgtagacacaccgcc
cgtcacaccacgagagtttgaacacccgaagccgggtggagtaaccaatttgagctagcc
5 gtcgaaggtgggacaaatgattgggggtgaagtcgtaacaaggtag

Sequence 1584

VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVLSSDCSLQLDYMKLES
LVIVDQHATVNTFPGLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVVTR*

Sequence 1585

Contig_0623_pos_15682_15359,

putative peptide of unknown function

gtgaaaaaagtgaaaaatacaaatagtaatatgcataaaaatgacaatagttacgtttta
15 tcagcattgagttatttaagtatcttttcgcacctgtcattttacctttatttgtagg
atacttgctgatgaaccaacatcagaacatggcaaaaaagcattcattaatcatattatg
acttgggttagtttttttataggtagattggcattttattttttctaaagaagtctttgat
aaacctttggatcatcaattattaattcatccgctcacttttcaacgtaagtcggttc
20 ggtcctccattcagtggttacctga

Sequence 1586

VKKVENTNSNMHKNDNSYVLSALSYSIFFAPVILPLFVWILADEPTSEHGKKAFINHIM
TWVSFFIGRLAFIFSKEVFDKPLDHQLLIFIRSLFNVSFRGPPFSVT*

Sequence 1587

Contig_0623_pos_11160_10618,

putative peptide of unknown function

atggttggtatgagagcgactgagagccatgatttaattcttgatgatgtttatgttcct
aatgaaaattttgtagaatcaaaaçgtgaatcaagacctaatggttggtcttctcatata
30 ccgagctgttatctaggtattgcacaagcagctcgtgactatgcagtagattttgcaaaa
aattatcgtcctaataatagattacaggtacgattgatagtttacctacagtgaacaaaat
ttagggaaaaatggaaagtttattactttctgcaagacattttctatggagtacagctaga
gggtatcaatcatatacacagaggatgcacaaatatggaatgaaacctcagcaagtaaaagt
gtggtaaatgaaccaaggtatagaaatcggttgatttagctatgagaatagttggagctaag
35 agtctagaaatgagcagacctcttcaacgggtactatagagatatacgtgctggattacat
aatccaccaatggaagatatggcttacactaatattgctaaaagtattacaacaaactt
taa

Sequence 1588

MVGMRAATESHDILDDVYVPNENFVESKRESRPNGWLLHIPSCYLGLIAQAARDYAVDFAK
40 NYRPNISITGTIDSLPTVQQNLGKMESLLLSARHFLWSTARGYQSYTEDAQIWNETSASKV
VVMNQIEIVDLAMRIVGAKSLEMSRPLQRYRDIRAGLHNPPMEDMAYTNIAXSITNKL
*

Sequence 1589

Contig_0623_pos_9215_8700,

putative peptide of unknown function

atgactgctatcatttgtatttttaggtatagtaccaagtgtacctctgccatttatgcc
gtccctattgtattacagaatataggaatctttttggcaggaattattttagggcgaaag
50 cttggtactactagtgttattgtctttttactattggttagctacaggtttgccagtgctt
tctggaggccgtgggggaattggcggtatttgcaggaccttcggcaggattcttattctta
tattcctgtttagcttactttataggcattattcgtgatgcataatttgcataaaattaat
ttcttagtgatttttatagctacactagttatcggtgtattaggattagatatattaggt
actatcattatgggctttattatacatatacctatctctaaggcatttatattatcattt
55 acatttatgccaggtgatatttaaggctattattgcaagtttaataggtgcagcaatt
ttaaatacttcacgtttcaagactcttattcaataa

Sequence 1590

MTAIIICILGIVPSVPLPFMPVPIVLQNIIGIFLAGIILGRKLGTTSVIVFLLLVATGLPVL

SGGRGGIGVFAGPSAGFLFLYPVVAYFIGIIRDAYLHKINFLVIFIATLVIGVLGLDILG
TIIMGFIIHIPISKAFILSFTFMPGDIIKAIIASLIGAAILNHSRFTLIQ*

Sequence 1591

5 Contig_0623_pos_8486_7908,
putative peptide of unknown function
atgcatctaaatcaatcatcctatcatctttcattttcattttccagctaataaaaaagata
gatgaagtattgttggaataaacgtgaactaggttttcagataggagtacttgagctc
tatgtcattgaagctaaagcggttaaaagagctctcccgcaaaagagacgtagatattcaa
10 cttgtatcaagcaataatatcaatgattaccttcatgtttatgatgcgtttgcacggcct
tttgggtgatagctatgccaaacatggttaacaacatatattatagctcatataaacttggac
gatattgaacgttttagttgcatatgttaaccatcaaccagttggaatagtcgatattata
atgacggataaaacaatagaaatagatggttttggggttttagaagaattccaacatcaa
ggtatcgggttctgaaatacaagcttacgttggacgtatggctaatagagcgacctgttatt
15 cttgttgcatagtgaaaagatactgctaaagatatgtatctaagacaaggatatgtatat
caaggttttaagtatcatatttttaaaagaaaatatttaa

Sequence 1592

MHLNQSSYHLSFSFPANEKIDEVLLKIRELGFQIGVLELYVIEAKALKELSRKRDVDIQ
20 LVSSNNINDYLHVYDAFARPFGDSYANMVQHIYSSYNLDDIERLVAYVNHQPVGIVDII
MTDKTIEIDGFGVLEEFQHQIGSEIQAYVGRMANERPVLVADGKDTAKDMYLRQGYVY
QGFKYHILKENI*

Sequence 1593

25 Contig_0623_pos_6439_5762,
is similar to (with p-value 5.0e-80)
>gp:gp|AB015195|AB015195_4 Staphylococcus aureus gene for Ly
tN and Eprh, complete cds. NID: g3767591.
atggacattattaaagaaatgaaaaagcaaatgttagttttacaacatactttgatgat
30 aactacccttctcttttgcaagaaatgtatgattatccttatgtgatattctacaaagga
aatccacagttcttttaatcattctcactcttttagctgtaattggctcacgtaatgccaca
caatatacaagtcaatctttaaactatcttttcttcatcttagacaattaaatatggcg
attgttctctggattagcgcgcggtgcatagatgttagcacatcaaaccgcacttaaatatc
ctattaccaactattggcgtacttggatttggcattgttatcattatcctaaagcaacc
35 ttaaatttaagaactaaagttgaaaggaatggcttagtgataagtgaatatccaccattt
tctcctataagtaagcataaatttctgaaagaaacaggccttataagtggtctgtccaga
ggggtacttataactgaggtgaagaaagaagtggttagtcaaatcactatcgattgtgct
ttagagcaaaatagaaatgtttatgttctacctggttcaatgttcaacaaaatgactaaa
ggtaatttaagaaggataaatgaaggtgctcaagttgttatagatgaaagtagtatatta
40 tatgattatctatttttag

Sequence 1594

MDIIKEMKKANVSFTTYFDDNYP SLCKEMYDYPYVIFYKGNPQFFNHSLSLAVIGSRNAT
QYTSQSLNYLFPSFRQLNMAIVSGLARGADSVAHQTALKYLLPTIGVLGFGHCYHYPKAT
45 LNLRTKVERNGLVISEYPPFSPISKHKFPERNRLISGLSRGVLITEAEERSGSQITIDCA
LEQNRNVYVLP GSMFNKMTKGNLRRINEGAQVVIDESSILYDYLF*

Sequence 1595

Contig_0623_pos_5489_3510,
50 is similar to (with p-value 0.0e+00)
>sp:sp|P39814|TOP1 BACSU DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME) (UNWISTING ENZYME) (SWIVELASE). >gp:gp|L27797|BACSMF_2 Bacillus subtilis (smf) gene, 3' end, DNA topoisomase gene, complete cds, (gid) gene, 5' end
55 . NID: g520751. >gp:gp|Z99112|BSUB0009_82 Bacillus subtilis complete genome (section 9 of 21): from 1598421 to 1807200. NID: g2633902. >gp:gp|AJ000975|BSYLQGCOD_6 Bacillus subtilis ylgg to codV gene region. NID: g2462964.
atgggacatgttcgtgacttaccagaagtcacaaatgggtgtcgacactgaagataactat

gaacccaaaatatattacaattcgtggcaaaaggtcctgtagttaaaagatttaaaaaaacat
 gcgaaaaaagcaaaaaaatatttttagcttagtgaccctgaccgtgaaggtgaagcgatt
 gcttggcattttatcaaaaaatttttagaattagaagatagcaagaaaatagagtagtattt
 aatgaaattacaaaagatgctgttaaagatagttttaagcatcctcgtgggtattgaaatg
 5 gatttagttgacgcgcaacaagcacgctgatttttagatagactcgttgggtataatatt
 tctccagttatttgaagaaagttaaaaaagggtcttctgctgggagagttcagtcagtt
 gctttacgttttagtcattgatcgtgaaaatgaaattcgttaattttaaacctgaagagtat
 tgggtccattgaaggtgaatttagatacaagaaatctaaatttacagctaaatttctacac
 tataaaaaataaaccttataagctaacaacaagacgatgttcaaaggattactgaagca
 10 ttaaatgggtgatcaatttgaaatcacaaatgtgaatcgttaaagaaaaaacagttatcct
 gctcatccatttactacatcaaccttacaacaagaagctgcacgtaaactaaattttaaa
 gcacgcgacaatgatgttagcacaacaattatacgaaggtattgacttaaaagcgtcaa
 ggtacagttaggtttaattacgtatatgctgaccgatttactcgtatctcaacttctgca
 aaatcagaagcgcagcaatatataaatgataaatatgggtgaacagtagctgtctcagcgt
 15 aaatcatcgggtaaacagggcgatcaagatgctcacgaagctattagacctactagtaca
 atgcgaactcctgatgacatgaaagcttttcttactagagatcaacaccgtctatcaaaa
 ttaatttgggaaagattttagcaagtcagatgggtccagctatttggatacagtagct
 ttagatgtaactcaaaacgacattaaatttagagctaattgggtcaactattaaatttaaa
 ggttttttagacactatatgtagaagcaaaagatgataaagagaatgataaagaaaaaag
 20 ctctcctcaactagataaaaggagataaggttaactgcgacaaagattgaaccggcacaacac
 ttacacaaacctcctcctcgttatactgaggcgcttttagttaaaacgcttgaggaactt
 aaaattggaagaccttcaacatatgctccaaccattgatacgattcaaaagcggaaactac
 gtcaagttagaaagtaaacgcttcatcccaactgaattaggagaaattgtttatgagcaa
 gttaaagaatacttcccagaaattattgatgtagaattcactgtaaacatggaaacatta
 25 cttgataaaaattgccgaaggtgacatgaattggcgtaaagtaaataggagacttctacaac
 agtttttaaacaaagattgtgaacgcgcagaatctgaaatggaaaagattgagattaaagac
 gagccagctggtgaagattgtgaagtcgtggttctccaatggttatataaatgggaaga
 tatggttaagtttatggcatgttcgaactttccggactgtcgtaacaccaaagcaattgtc
 aaaacgattgggtgtcacatgtccgaagtgtaatgaaggagatgtcgtagaacgtaaatca
 30 aagaaaaatagaattttctatgggtgttcttagatatccagaatgtgattttatttcttg
 gataaacctgttggaagagattgtcctaagtgatcattaccttgtgaacaagaaaaaa
 ggtaaaaagtagtcaagttgtgtgctccaactgtgattatgaagaagaagttcaaaaatag

35 Sequence 1596

MGHVRDLPRSQMGVDTEDNYEPKYITIRGKGPVVKDLKKHAKKAKKI FLASDPDREGEAI
 AWHLSKILELEDSKENRVVFNEITKDAVKDSFKHPRGIEMDLVDAQQARRILDRLVGYNI
 SPVLWKKVKKGLSAGRVQSVALRLVIDRENEIRNFKPEEYWSIEGEFRYKKSFTAKFLH
 YKNKPYKLNNKDDVQRITREALNGDQFEITNVNRKEKTRYPAHPFTTSTLQGEAARKLNF
 40 ARKTMMLAQQLYEGIDLKRQGTVGLITYMRTDSTRISTSASEAQYINDKYGEQYVSQR
 KSSGKQGDQDAHEAIRPTSTMRTPDMDKAFLTRDQHRLYKLIWERFVASQMAPAILDTVA
 LDVTQNDIKFRANGQTIKFKGFMTLYVEAKDDKENDKENKLPQLDKGDKVTATKIEPAQH
 FTQPPPRYTEARLVKTLEELKIGRPSTYAPTIDTIQKRNYYKLESKRFIPTELGEIVYEQ
 VKEYFPEIIDVEFTVNMETLLDKIAEGDMNWRKVIGDFYNSFKQDVERAESEMEKIEIKD
 45 EPAGEDCEVCGSPMVIKMGYKFKMACSNFPDCRNTKAIVKTIGVTCPCNEGDVVERKS
 KKNRIFYGCSRYPECDFISWDKPVGRDCPKCHHYLVNKKKGKSSQVVCNSCDYEEVQK*

Sequence 1597

50 Contig_0623_pos_3509_2178,
 is similar to (with p-value 0.0e+00)
 >gp:gp|Z99112|BSUB0009_83 Bacillus subtilis complete genome
 (section 9 of 21): from 1598421 to 1807200. NID: g2633902. >
 gp:gp|AJ000975|BSYLQGCOD_7 Bacillus subtilis ylgg to codV ge
 55 ne region. NID: g2462964.

atgagtgatttttaggaggggcaaaatgacgcaaaaagtaaacggtttaggagctggttta
 gctggctctgaagctgcatatcaattagctcaacgtggaattaaagtgaatttaattgag
 atgctgccagttaaacagacaccggcgaccatacagataaatttgcgtgaattggtatgt
 tcaaatcattgagaggttaatgcacttacaaatgctgttggtgttcttaagaggaaatg

agacatttagactcggttaattatcacatcagcagataaagcacgtgtgccagcgggtggt
 gcttttagcagtggatagacatgattttgctggctatattacagataccttaagaaaccac
 cctaacatcactgtattaaatgaagaagttaatcatataaccagaagggtatacgtattatt
 gcaactggccctctaactactgagcatttagctcaagaaattggtgatattactggtaaa
 5 gatcaattgtatttttacgatgctgccgcaccaataatagaaaaagattcaattaatatg
 gataaagtatatgttgaatcacgttatgataaagggtgaagcagcgtatcttaattgtcct
 atgactgaagaagaggtttaaccggttttatgatgcagtattagaagctgaagttgcacca
 gtcaatgagtttgaaaaagaaaaatattttgaagggtgatgccttttgaagtcattggct
 10 gaaagagggcgaaaaaactttgttatttgggtccgatgaaacctgttgacttgaagatcct
 aagactgggaacgccttattgcagttgttcaattaagacaagatgatgcagctggaaca
 ttatataatattgttggctttcaaacacattttaaattggggtgcgcaaaaagaagtcatt
 cgtttaattccaggattagaaaaatgttgatattgtgaagatatggtgtgatgcaccgaaat
 acctttattaattcacctgatgttttaaacgaaaaatgaattaaaaggacatgataat
 ttatattttgctggacaaatgactggcgttgaagggttatgttgaaagtgtgccagtgga
 15 ttagttgcaggtattaatcttgcgcataaaattttagacaaaggggaagttattttccct
 agagagacaatgataggttagtattggcttactacatatcacatgccaaaaatgagaagaat
 tttcaacctatgaatgccaattttgggtcttttaccatctctcgaaaaacgtattaaagat
 aaaaaagaaagatatgaaacacaaagccaaaagagcgtagagtatttagataattacaaa
 caaacgctgtaa

20 Sequence 1598
 MSDFRRGKMTQKVN VVGAGLAGSEAA YQLAQRGIKVNLIEMRPVKQTPAHHTDKFAELVC
 SNSLRGNALTNVAVGLKEEMRHLDSLIIITSADKARVPAGGALAVDRHDFAGYITDTRLNH
 PNITVLNEEVNHIPEGYIIATGPLTTEHLAQEIVDITGKDQLYFYDAAPIIEKDSINM
 25 DKVYLKSR YDKGEAA YLNC PMTEEEFNRFYDAVLEAEVAPVNEFEKEKYFEGCMPFEVMA
 ERGRKTLLEFGPMKPVGLEDPKTGKRYPYAVVQLRQDDAAGTLYNIVGFQTHLKWGAQKEVI
 RLIPGLENVDIVRYGVMHRNTFINS PDVLNEKEYELKGHNDLYFAGQMTGVEGYVESAAASG
 LVAGINLAHKILDKGEVIFPRETMIGSMAYYISHAKNEKNFQPMNANFGLLP SLEKRIKD
 KKERYETQAKRALEYLDNYKQTL*

30 Sequence 1599
 Contig_0623_pos_1923_967,
 is similar to (with p-value 4.0e-47)
 >sp:sp|P39776|CODV_BACSU PROBABLE INTEGRASE/RECOMBINASE CODV
 35 . >gp:gp|U13634|BSU13634_2 Bacillus subtilis JH642 dipeptide
 permease operon regulators, codV, codW, codX, and codY gene
 s, complete cds. NID: g535347. >gp:gp|Z99112|BSUB0009_84 Bac
 illus subtilis complete genome (section 9 of 21): from 15984
 21 to 1807200. NID: g2633902.
 40 atgttaaagggttgaagaaaacttttcagagtatacgtttaaactttatcatgatgattta
 gttcaatttaacaactttttagaaagagaacatttacaacttgagacttttgaatataaa
 gatgctagaaactattttgcttttttatattctaatacaattaaaaagaactacggtgtca
 agaaagatatcaactttacgtaccttctatgaattttggatgactcaagataattcaatt
 45 attaatccctttgttcaactagtgcacctaataaaagagaagattttacctcaattcttt
 tatgaagaagaaatggaagcactttttcaactgtagagcatgataataaaaaaggcata
 cgagacaaagtattattgaattgttatatgcaacaggaatacgtgtgtctgaattaata
 aatattaaactaaaagatatagatatgaacttaccaggtgtaaaagttttaggtaaagga
 aataaggaaaaggtttatcccttttggagagttctgtagacagagtatagaaagatactta
 gaagaattccaacctaaacaattagccaatcatgattatttaattgtaaataatgaaagggt
 50 gatcctatcacccgaaagaggagtaagatatgtacttaattgatgtcggttaaaagaaccgct
 ggcgtcaatgacatacatcctctattttatgattatagcgacgatagggtgaaatggtatat
 tctccaattcttgaagaaaatcgttttaaaatgggttccttctcataaaagaggacatat
 tcagcagtgcagcttttaggatttaacctagctgaattacttgcaagatttggaattata
 ttaggagtgtttttaactttcaatggagatggggatctatatgtttgttttattattacta
 55 ggtggcatgtcactttacattgcagtgagtcgttttaataatacaaaattcacataaa

Sequence 1600
 MLKVERNFSEYTLKSYHDDL VQFNNFLEREHLQLETFFEYKDARNYLAFLYSNQLKRRTVS
 RKISTLRTFYEFWMTQDNSIINPFVQLVHPKKEKYL PQFFYEEEMEALFQTVEHDNKKGI

RDKVIIELLYATGIRVSELINIKLKDIDMNLPGVKVLGKGNKERFIPFGEFCRQSIERYL
EEFQPKQLANHDYLIIVNMKGDPITERGVRYVLNDVVKRTAGVNDIHPLFMIIATIGEMVY
SPILEENRFKMPVPSHKRGTYSAVHALGFNLAELLARFGIILGVFLTSMEMGIYMFVLLLL
GGMSLYIAVSRFNNTNSQ*

5

Sequence 1601

Contig_0623_pos_0_694,

putative peptide of unknown function

gtgacatactgggttaaagtttaagtcgcgacattgaagttacgtagattaatgtatgcatta
10 ttagatgtcgttcaaagtcacacctgtgttgcgtacacagtttgtgacagatgattttaat
caactcaagataaaatttaagagattttttccatttattgaaattaaagaagttaatgaa
atgtcgcgaaagcatagatttagaagcattctttacacgtaatttaaattcctaccatttc
aatcaattacctctgtttaattttaagatatatcaatttctagatggcgccctacctactt
ttagattttccacgctactatttttaagaaagtcatttaactccatttttacaacaatta
15 aatattgcttatacccactctttaaaaagtgaaatagtagtctcggatttttataattgg
attaaagaaatgaatcaaaagatggatcaaaatcaagttgtgtgtccatcaaagcacttc
aacgtattgaatgcagacgggtgataattacgcttacataacctgtaaagaatacatctgaa
aagaaaaaatgtgttctttgcatgcagaactaccatctttagacattgatgcgtggatt
gtaagtatttacttagcgcatctttataagtcagtcctctgatgtgacgttaggcacac
20 catttttcgatagataataaaaaatactgagaata

Sequence 1602

VTYWVKLSRDIELRRLMYALLDVQSQPVLRTQFVTDDFNQLKINLRDFFPFIEIKEVNE
MSQSIDLEAFFTRNLNSYHFNLPLFNFKIYQFLDGAYLLLDHFHATIFNESQLTPFLQQL
25 NIAYTHSLKSEYSISDFYNWIKEMNQKMDQNQVVCPSKHFVNLADGDNYAYIPVKNTSE
KKKMCSLHAELPSLDIDAWIVSIYLAHHFISQSSDVTLGIHFSDNKNTENX

Sequence 1603

Contig_0625_pos_5621_6001,

putative peptide of unknown function

gtgatttcatctactggttgttttgttatttttctgttgggttcaccttcgccaaactttt
30 tccccgttaaatgggttcttagttgttgggtgtgtaattgttttctcctggttcacct
ttctgtttaacacgctcttcacctgggttttaaatcaggattgaattcacgtttcttgctg
aatggaatttcttccgttgacgtaatcggatctccatcaactggaccatattttgtcaca
35 tcatccactggtggtgtgactacttcgcctgtatcaggatttttaactcctggtttacct
ggaacgtcctcttggtacctttcggtgcatttggatcaaattcatccttatggcctggc
ttgatttcttcgccaccataa

Sequence 1604

40 VISSSTGCFVIFSVGSPSPFSPVNGFLVVGVVIVFVPGSPFCLTRSSPGFKSGLNSRFLS
NGISSVDVIGSPSTGPYFVTSSSTGGVTTSPVSGFLTPLGLPGTSSWLPPGAFGSNSSLWPG
LISSPP*

Sequence 1605

45 Contig_0625_pos_6773_7471,

putative peptide of unknown function

gtgatttcatctactggttgttttgttacttttctgttgggttcaccttcgccaaactttt
tccccgttaaatgggttcttagttgttgggtgtgtaattgttttctcctggttcacct
ttctgtttaacgcgctctttacctgggttttaaatcaggattgaattcacgtttcttgctg
50 aatggaatttcttccgttgacgtgatcggatctccatcaactggaccatattttgtcaca
tcatccactggtggtgtgactacttcgcctgtatcagggtttttaaactcctggtttacct
ggaacgtcctcttggtacctttcggtgcatttggatcaaattcatccttatggcctggc
ttgatttcttcgccaccatattctgtgatttcatctactggttgttttgttatttttct
gttgggttcaccttcgccaaacttttccccctgttaattgggttcttagttgttgggtgtgta
55 attgttttctcctggttcaccttttgtttaaacgcgctcttcacctgggttttaaatca
ggattgaattcacgtttctgtcgaatggaatttcttccgttgacgtgatcggatctcca
tcaactggaccatattttgtcacatcatccacaggtggagtaactacttcgcctgtatca
ggatttttaacccccggcttacctggttgcgttgtttga

Sequence 1606

VISSTGCFVTFVSGSPSPTFSPVNGFLVVGVVIVFVPGSPFCLTRSLPGFKSGLNSRFLS
NGISSVDVIGSPSTGPYFVTSSSTGGVTTSPVSGFLTPLPGTSSWLPFGAFGSNSSLWPG
LISSPPYVISSTGCFVIFSVGSPSPTFSPVNGFLVVGVVIVFVPGSPFCLTRSSPGFKS
5 GLNSRFLSNGISSVDVIGSPSTGPYFVTSSSTGGVTTSPVSGFLTPLPGCVV*

Sequence 1607

Contig_0625_pos_8861_7932,

is similar to (with p-value 0.0e+00)

10 >sp:sp|Q05615|AROA_STAAU 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTR
ANSFERASE (EC 2.5.1.19) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE
SYNTHASE) (EPSP SYNTHASE). >gp:gp|L05004|STAAROA_2 Staphyloc
occus aureus dehydroquinase synthase (aroB) gene, 3' end cds
; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene,
15 complete cds; ORF3, complete cds. NID: gl52954.
atggatagagtgatgaaacctctgttaaaaatgaatgctaataatctggaattgataat
aattacacaccacttataattaagccttcaactattaaaggtataaattatcaaattgaa
gttgcaagtgcacaagttaaaagtgtctatcttacttgctagcctatcttcaaaagaagcc
acgacacctacagaatttgatgtaagttagaaatcatacagaacattgtttgcacacttt
20 aacattcctatatcaattcaaggtaaaacaatccaaaccataccttatgcaattgaaat
atacagcctagagattttcatgttccaggagatatctcatctgctgcatttttcatagtc
gcagccctaattacgcccgttagtgacattacaattcataatgttggcattaatcccact
cgttcaggtatcatagatatcggttaacaaatgggaggaaacattgaattaagtaattgta
agcaaggggtgcagaaccaactgcatcaatacatgtaaaatatacaccgaacttgaaatgct
25 gttacaattaaaggcgacttagttccaagggtatcgatgaattaccagttattgcacta
ctttgcacacaagcttcaaatcttctgtattatcaaaaacgcggaagaattaaaagtgaag
gaaacaaatcgtagatactactcgcagacatgctaaatttacttggttttaatttaca
cccacacatgatggacttattatacatccatcagagtttagatcaaatgcaactgtagat
agtcaaacagatcatcgtagaatgatgttagctgtagcttactacttagttcagaa
30 ccactcaaaatagagcaatttgacgctgttaattgtttccttcccaggtttttacctaaa
ctaaagcttttagaaaatgagggaaaataa

Sequence 1608

MDRVMKPLLMKNANISGIDNNYTPLIIPSTIKGINYQMEVASAQVKSAILLASLFSKEA
35 TTLTEFDVSRNHTETLFAHFNIPISIQGKTIQTIPYAIEHIQPRDFHVPDIISSAAFFIV
AALITPGSDITIHNVGINPTRSGIIDIVKQMGGNIELSNVSKGAEP TASIHVKYTPNLNA
VTIKGDLVPRAIIDELPVIALLLCTQASNSCIIKNAEELKVKETNRIDTTADMLNLLGFNLQ
PTHDGLIHPSEFRSNATVDSQTDHRIGMMLAVASLLSSEPLKIEQF DAVNVVSFFGFLPK
40 LKLENEGK*

Sequence 1609

Contig_0627_pos_595_1704,

putative peptide of unknown function

atggtaaacagtaaatgatattgtttctattgttattagtgatattacacgtccaacgccc
45 aaccatattcttgtacctttactaattgaggaattaaatcatgttcctcgtgagaatttc
gtaattattaatggtacagggactcatcgagatcaaacgcgagatgaattgattcaaatg
ttagtggaagatattgtaattcagtaaaaaatcggtcacaatcattgctcagaaaaagaa
agtctagctaaagtgggacacagtcataatggatgtgatgtttatttaacaaaacatat
gtagaatccgatttttaaaattgtaacaggttttattgaaccacactttttcgccgatttt
50 tcaggtggacctaaagggataatgcctggaattgcaggtttagaacaattcaaacattt
cataatgcaaaaatgattggcgatccgagatcaacgtggggaatttagaagacaatcca
gttcaagatatggcacgggaagttaaccgtatgtgtaaacctgactttttacttaattgtt
gcattgaataaaaagtaaaagaaattactgcagcatttgctggtgaaatcttagatacacac
aaagaaggatgcgcataatgtaaaagatcatgcaatgtttaaatgtgagcaacgctttgat
55 attgttatcgcatcaaatctctggctatccttttagatcaaaaatttatatcaaacagttaaa
gggatgagtgacgcagtaaaagtgttaaaaaagacggtcatattattatggtatctgag
tgtgcagatggcttttctgatcatggtaagtttgccgaaattttcaaaatggcagacaca
cctcaaggatttttagaacttattcacaatccaaactttaagggaagtgaccaatggcaa
gtacaaaaacaagcaagtattcaaaccttttccaatgtgcatgtttattcagaacttact

gaccaacaacttaagactcgatgttaatcccaaccttaacattgaacatacaatacaa
gaattagaacatcgatatggccgtaaattaaccattggtgttatgccacaaggctccttta
acaataccgtacgtagaagataaagaataa

5 Sequence 1610

MVNSNDIVSIVISDITRPTPNHILVPLLEELNHVPRENFVIINGTGTHRDQTRDELIIQM
LGEDIVNSVKIVHNCSEKESLAKVGHSQYGCDEVYLNKAYVESDFKIVTGFIEPHFFAGF
SGGPKGIMPGIAGLETIQTFHNAKMIGDPRSTWGNLEDNPVQDMAREVNRMCKPDFLLNV
ALNKSKEITAAFAAGEILDTHKEGCAYVKDHAMFKCEQRFDIVIASNSGYPLDONLYQTVK
10 GMSAASKVVKKDGHIIIMVSECADGFPDHGKFAEIFKMADTPQGILELIHNPNFKEVDQWQ
VOKQASIQTFANVHVYSELTDQQLKDSMLIPTSNIEHTIQELEHRYGRKLTIGVMPQGFL
TIPYVEDKE*

Sequence 1611

15 Contig_0627_pos_1748_2551,
is similar to (with p-value 1.0e-44)
>sp:sp|P73846|YH17_SYNY3_HYPOTHETICAL_30.2_KD_PROTEIN_SLR171
7. >gp:gp|D90910|D90910_32_Synechocystis_sp._PCC6803_complet
e_genome_12/27_1430419-1576592. NID: g1652956.
20 atgaaattagacgcactattgaaagacatgcagagtgtagtaattgccttctcaggtgga
gtagatagtagcttggtactgaaaaagcgattgatatttttaggtgttaactatgttaaa
cctgtttagtagtaaaatcagaattatttagaaatgaagagtttgaactagcgcttaaaactt
ggacaaagtctaggtgttgaaagtattagaaactgaaatgtctgaacttcaagatgcgaat
atcggttaaaaatcgcctgaaagttggtactatagcaagcgcttgatgtatagtcaactt
25 gagaatattaagaataaaactaggatttaattatgtgctagatggtatgattatggatgac
ttagatgattttcgctcccgattaaaagcaagagacgactttggtgttcgtagcggttta
caagaagcaaaactatataaatctgaagtttagagaattaagtcacatcaacatgacttgcct
gtatggaataaaccagccttatgtagtctagcatcaagaataccttatggtgaggaatta
agttttacaaaaggttaacaaggtcaacgaagcagaaaaattcattttaagcctaggtatt
30 aaccacgtacgagtacgctatcatcacaacatagcacgcattgaagtaacagaagatcaa
ttaaataatcttcttaattgaaagacagtatcatattacatttgaagaattaggattt
gactatgtaacaatggatttagaagcgtatcgtaggttagtatgaatgaaatcattgat
accaaactcacaagttttaataa

35 Sequence 1612

MKLDALLKDMQSVVIAFSGGVDSSLLKKAIDILGVNYVKPVVVKSELFNNEEFELALKL
GQSLGVEVLETEMSELQDANIVKNTPESWYYSKRLMYSQLENIKNKLGFNVLDMIMDD
LDDFRPGLKARDDFGVRSVLQEAELYKSEVRELHQHDLPVWNKPALCSLASRIPYGEEL
SFTKVNKNVNAEKFILSLGINHVVRVYHNNIARIEVTEDQLNNLKLKDSIILHLKELGF
40 DYVTMDLEGYRTGSMNEIIDTKSTSFK*

Sequence 1613

Contig_0627_pos_2566_3342,
is similar to (with p-value 8.0e-40)
45 >sp:sp|Q57629|Y165_METJA_HYPOTHETICAL_PROTEIN_MJ0165. >gp:gp
|U67473|U67473_9_Methanococcus_jannaschii_section_15_of_150
of_the_complete_genome. NID: g2826256.
atgagccatagttataattctatagaagaggtgctcaaagctgtaaaatcaaatacaacta
tctattaatgatgctaaagcccaactcagtcattatgacgaattgggctttgctaaaatt
50 gacttacatagagcacagcgtcaaggatttcccgaagttatctttgggcaaggaaaaaca
aaagaacaaaatcactaaaatcatctctagtttgatattttcataatgaagttattctagt
acacgtgttgatgaaatgaaagcaaaaatacattttacaacattatccaaacttggaatat
catcaaaactgcacagtttaattagcactccactaaaagatataccacaatctaataactat
gtttctgtactttgtgctggaacttctgatttacctattgcagaagaagctgcattaacc
55 gctgaaatcatgggagtaagtgtaaaacgattttatgatgtcgggggttcaggtattcat
cgcttattatccaacattcatgatatacgcagagggaaagtttctatcgttatagctgga
atggaaggcgcttttagcaagtgttggtgaggattagtcaccaccctgtatagcagta
ccaacgagtgtaggttatggagcaaaacttgaatgggggttaccaccctattatcaatgata
aatagttgcgcacccggaaccagcgtatttaaatatcaataatggatttgggtggcggttac

aacgctgcacagattattcatatgctagaaaataaagagagtggaggtatctttatga

Sequence 1614

MSHSYNSIEEVLKAVKSNQLSINDAKAQLSHYDELGFAKIDLHRAQRQGFPEVIFGQGKT
5 KEQITKIISLIFHNEVILVTRVDEMKAQYILQHYPNLEYHQTAQLISTPLKDIPQSKYY
VSVLCAGTSDLPFAEEAALTAEMGVSVKRFYDVGVSIGIHRLLSNIHDIRRGKVSIVIAG
MEGALASVVGGLVNHVPYAVPTSVGYGANLNGVTLLSMINSCAPGTSVLNINNGFGGGY
NAAQIIHMLENKESEVSL*

10 Sequence 1615

Contig_0627_pos_3387_3701,

putative peptide of unknown function

atgctacttttctgcttttagttgatttaggagcaaaccctgaagacattgaatcagaacta
aaaaaattaccttttagatcaatttaagctacattttcaaaaaagagtaaaacaaggatt
15 catgcaatgacattaacattgatgttaaagaagcaaatcatcatcgtcacgttaaatgat
atattttaaagtatagatgacagtacacttccggaaagggttaaataatcgacagtaagaaa
atttttgaaatcattgggtcaagcagaagctaaaattcatggcatgtcttcaagcatgttg
ttaactttactttga

20 Sequence 1616

MLLSALVDLGANPEDISELKKPLDQFKLHFQKRVKQGIHAMTLNIDVKEANHHRHVND
IFKMIDSTLPERVKYRSKKIFEIIGQAEAKIHGMSSSMLLTLL*

Sequence 1617

25 Contig_0627_pos_5833_7239,

putative peptide of unknown function

atggcaattctaaccaccattatcaactacggattggcatgcatataaagttaatctaagt
caatatttgactcaagaaaatggtcgttatttaggacattttattgaatgggttgccgta
cataataacataataagagctttaatatatgcgataaacttcgttttttagttatctattta
30 gttgcttatatgggtcaattacatacgaatcgtattttatttttttaggtttgtgtta
atgggttactgtacctaatacaatttatagcgaaaacttacgggtgggttactggatttttt
agttatatacctgctacagtcctatcacctttttattctttttacggtagttaaaaagatt
gagtcgcacgatacagtttctgaaatgcaattatgggtattttttattagtaagttgttt
ggacaattcttcttggaagaatcttccatcgctaataagcttaatttttaataaggaatg
35 gtagtctatttcttggtaaaaaagactcagttatttcttaattgtaggatttatgctt
agttgtataggttaacattataatgtttttaaacttcaattattttttaattaaggatgga
ttaaatacgcattattcaatttccgatagtcaggaatgatacataaagcaggtgtgacg
ttatttaagcttgaccagaatatatgtttattaatcaaatgattattcttaccgtgata
tcaatagtaagtatagttttacttaagcaaaaataaaagcctgaagcatatgagagtttat
40 attaaaaataccactactcttaggtttaattactttacctatttataagatcttcgtttac
aatcaatttcattttgaattatataaagcttcattttctatagccgttttgatacaacg
atttgcttcattttacatgataagtgatgatacgttggtgtttaaaatgatacagcaaaga
tacataagaatgattgtgatggggagttttatagctatggcttcacgtgttttgccactt
ttatttgtgacgcctataagttatagaaattttttattttatttatactttatggatcgtg
45 atattactttgtttaattcagcaatgtgatgtgctatttaacaacttgaacatataatt
aaaatatttgcgattatcatcagcatcattatgatgattggatttacttttatacatatt
agtagtgtgcacagaatagacttcattaaagaacaaataacacaacatcatcgctatcag
aaaataacattggaaagattaccatttgagcgatatactcatatgactacaccaaagtcg
aaggaacaacttcaagatttcaaacactattatgatttgcccaaagacatcacatttaaa
50 gtagtcccatatggtacaaaacaataa

Sequence 1618

MAILTPLSTTDWHAYKVNLSQYLTOENGRYLGHLFEWVAVHNNIIRALIYAITSFVLVIYL
VAYMVQLHTNRIYFILSFVLMVTPNTIYSETYGWFTGFFSYIPATVLSLFILFTVVKKI
55 ESHDTVSEMQLWVFLVSLFGQFFLENLSIANSLIILIGMVVYFFVKRSLSYFLIVGFMJ
SCIGNIIMFLNFNYFLIKDGLNTHYSISDSHGMINKAGVTLFKLVPEYMFINQMIILTVI
SIVSIVLLKQNKSLKHMRYIKIPLLLGLITLPIYKIFVYNQFHFELYKASFSAVLNNTT
ICFIYMISVIYVVFKMIQQRIRMIVMGSFIAMASSVLPLLFVTPISYRNFYFIYTLWIV
ILLCLIQQCDVLFKQLEHIKIFAIISIIMMIGFTFIHISSVHRIDFIKEQITQHHRQY

KITLERLPFFERYTHMTTPKSKEQLQDFKHYYDLPKDITFKVVPYGTKQ*

Sequence 1619

Contig_0627_pos_7260_7646,
5 putative peptide of unknown function
atgaagttaaccgaatacattatgagattataaagtttatcatagttggtggaattaat
acctttaactactatataacatacttatttttgttaaagggtgttacatgtgaattatatg
gttagtcacattgttggatttattgttaagttttattatttcattattttaattgttat
10 tttgtatataaagtaaaacctacaatagaaaagtttttaagatttcctatcactcagata
gttaatatggtaatgcaaacgttattattatatatttcgtaaagggttgaatatcgct
tcagaaaattgcaccttttgcgggtctaatacattacaatccagtgacattcatactttct
aagtggttacttagagataaaagtttaa

Sequence 1620

15 MKLTRIHYEIIFKFIIVGGINTFNYYITYLFLLLKVLHVNYMVSHIVGFIVSFIISYYLNCY
FVYKVKPTIEKFLRFPITQIVNMVMQTLLEYIFVKWLNIASEIAPFAGLIITIPVTFILS
KWLLRDKV*

Sequence 1621

20 Contig_0627_pos_4306_3668,
putative peptide of unknown function
atgaaaagAACAGATAAATATAGAGATTcatacaaatatgatgaccaatatcaaaatcat
cgtaaacgttcagaagaagatatgtatcgacaacatcaagagtcaccaacagagagcaaat
tcaaactcgTCAACACAAGTGAAAATGATAGAGAGTATGAAAATCATCCTGAACGTTAT
25 tcaaatggaaagagactatcgacgtgagcagcaattggaagaagaaaatgaaaaatcaagc
aaaactaaaaaatggctgattgcaatcatagttattttactcattattgttagctatcttt
atcacacgtgcaattatcaatcataataatgataaagtaagtaatgaccctaacgtttca
caaaactataaaaaagaagttgaaaatcaaaacgacgacattaatcgacaagttgattca
gccaaaagcgatataaaaaataaaaaggacaccaatcccaaatgataaaactacaaaat
30 caaattgatcaattaaaacaaaatgaagaaactaatgCGGATTCTAAATTCACAAAATTT
tatcaaaaccaaactcgacaaactgaaaaatgcaaataacgctcaacttaataacgaaaaat
caaagtaaaagttaaacaacatgcttgaagacatgccatga

Sequence 1622

35 MKRTDKYRDSYKYDDQYQNHKRSEEDMYRQHQESQQRANSNRATQSENDREYENHPERY
YNGRDYRREQQLEENEKSSKTKKWLIAIIVILLIIVAFITRAIINHNNDKVSNDPNVS
QNYKKEVENQNDIDINRQVDSAKSDIKNKKDTQSQIDKLQNDIDQLKQNEETNADSKFTKF
YQNQIDKLKNANNAQLNENQSKVNNMLEDMP*

40 Sequence 1623

Contig_0629_pos_3864_2371,
putative peptide of unknown function
atgatttcatatgcgtgtagcaataataggtatgggaacagctgggtgtaagtgtgttacgc
caactcgttaaACATGAAAACttttctcaattaaaagtagatgtatatgacgatgttaga
45 aatatgggccaaggtgttccatttcaaaatgatagtagcgaactacttattaacatgcc
tcaaaatccatgagcttaaatcttgatgatgatcaagagttttggaagtggatcaaaat
cagacggaatttaatttttagtaatcctcaatatttgcttagatttgtatttggtcattat
atgaagtccttattttatcttattataatgaccaatttgataatttaactattatcaatgat
aaagtacaagaaatttttacacaatccgatgttgatgacacagattttaaataatcatgta
50 tgtacatgtgatgatgaaaaagaatggcgtgaatacgattatttatttttaacttttgg
acttttagttaccatgacccttatgatttgaagggaactaaaggctatatacaaacgccg
tatectacatatcacacttgataatgttaaagattcagatcgaatcgatgattattgga
acaggttggctagtttagatgctgtgagatattgttagctgcacatcatccatctttaccc
attactatgacaagtcgttctgcagcattgccaaagtgttagagggaaaatgactaaaatt
cagtttacgcatttaactaaatcacgatttaattgaatttatgaaaaatcactttggtaat
55 gtaccattagaaaaagtagtttcattatttttaaagaatgtgaagattatggaatagat
tttaaaaaacttattttatcgtagaacggaaaccatgtcaaaagacttggagtatgattta
aatcatgaagaagaaatggggatattccaaagtatcattgaacatttaaaagaaaattla
aactggatttggaaatagtttgagcgttaaagatcaagaaacttttaatcgtaaatacact

aaaattattcagttaaattctaataccaatgcctcctagaacagctcgtttacttatcaag
 ttaatacagaataatgaacttgtcattaaaaaagggtagaagacatagtcataaaaaat
 aatcaatttatgttggaagtataacgacactacgcaaaattatgagttggttgacatcggt
 5 attaatgcaacgggctctaaaaacacatctttctcaattagatgaggatgatcaattaatt
 ttaaacttagaaaatagacaaattgttcaacgtcatcctatgggtggcattcaaattatc
 ccagaaacaaatcaagtcataagccccagatatggaaccttaaaaaatgtgattgcaatt
 ggacaaatgaccaacgggtgtcaataaaacttagaaatggcgtaaagatgattgttaataca
 gttgttgatacagtatctcaattatatataacacaggaaaatagaaataagtaa

10 Sequence 1624

MIYMRVAIIGMGTAGVSVLRQLVKHENFSQLKVDVYDDVRNMGQGVFPQNDSSSELLINMP
 SKMSLNLDDQEFWKWYQNQTEFNFSNPQYLPRFVFGHYMKSYLSYNDQFDNLTIIND
 KVQEIFTQSDVDDTLKYHVCTCDEKEWREYDYLFTFGTFSYHDPYDLKGTGKIQT
 YPTYHTLDNVKDSRIVIIIGTGLASLDAVRYVAHHPSLPITMTRSRAALPSVRGKMTKI
 15 QFTHLTKSRFNGIMKNHFGNVPLEKVVSLFKECEDYGIDFKLIYRRTGNHVKDLEYDL
 NHEEEMGIFQSIIHLKENLNIWNSLSVKDQETFNRYTKIQLNSNPMPPTARLLIK
 LIQNNELVIKKGLEDIVHKNNQFMLKYNDTTQNYELFDIVINATGSKTHLSQLDEDDQLI
 LNLENRQIVQRHPMGGIQUIPETNQVISPRYGLKNVIAIGQMTNGVKNLNGVKMIVNQ
 VVDTVSQLYITQENRNK*

20

Sequence 1625

Contig_0629_pos_1922_891,

is similar to (with p-value 3.0e-59)

>sp:sp|P17618|RIBG_BACSU RIBOFLAVIN-SPECIFIC DEAMINASE (EC,3
 25 .5.4.-). >pir:pir|S45543|PN0100 ribG protein - Bacillus subt
 ilis >gp:gp|L09228|BACDIA_10 Bacillus subtilis spoVA to serA
 region. NID: g410114. >gp:gp|X51510|BSRIB_2 B.subtilis ribo
 flavin biosynthesis operon ribG, ribB, ribA, ribH, and ribT
 genes. NID: g40083. >gp:gp|Z99116|BSUB0013_40 Bacillus subti
 30 lis complete genome (section 13 of 21): from 2395261 to 2613
 730. NID: g2634723.

atggatgatgctattcaactagcaaaaatggtaaatggacaaacaggtgttaatccacca
 gtaggatccgttggttgtaaaaacggtaggattgtaggttttaggtgcacatttaaaaag
 ggagataaacatgcccgaagtacaagctattgaaatggcaggtttaaatacccaaggtgct
 35 accatatacgttttcattagaaccttgacacacaccatggttcaacaccaccttggtgcat
 aaaatcattgaagcgggcatatctaaggtcatctatgctgttaaagatactacttttagta
 agtaaggggtgacgagattctgagagaagctggtatagaggttgaatttcaatataatgaa
 aatgcagctgcattataaccgtgacttttttactgctaaaagaaacgaagttccagaagta
 actgtaaaaggtctcatctagtcttagatggtaaacagcaacagactttaatgaaagtaag
 40 tggataacaaacaaagaagttaaagaagatgtttatcaattaagacatgagcatgatgca
 gttattactgggcgtagaaccattgaagcagacaatccattgtatacaaccagggttcct
 gatggaaagcatccgattcgagttattctttctaagaaaggtcaactcgattttaatcaa
 caaatatttaaagatactgcatcgagatatggatttacactgaaaatgaaaaattaaaa
 acaataaaaagtttttattaaaataataaatattagtaattgtgatacaacgacaatatta
 45 caagacttatatacaagaggtattgggaaactgctagtcgaggcaggcccaaatattaca
 tctcaatttctccaatccaaacatctaaatgaactcattttatatatagccccgaaatta
 attggtgggtctggcaaacatcaattttataagactgacgaggtcattgatttgctgaa
 gcaactcaatttgaaattgttgattccaagttaattaatcaaaatttaaaattgaaatta
 cgaaagaagtga

50

Sequence 1626

MDDAIQLAKMVNGQTGVNPPVGSVVVKNRIVGLGAHLKKGDKHAEVQAIEMAGLNTQGA
 TIYVSLEPCTHHGSTPPCVHKIIEAGISKVIYAVKDTTLVSKGDEILREAGIEVEFQYNE
 NAAALYRDFFTAKRNEVPEVTVKVSSSLDGKQATDFNESKWI TNKEVKEDVYQLRHEHDA
 55 VITGRRITIEADNPLYTTRVPDGKHPIRVILSKKGQLDFNQI FKDTASEIWIYTENEKLLK
 TNKSFIKIINISNCDTTTTILQDLYQRGIGKLLVEAGPNITSQFLQSKHLNELILYIAPKL
 IGGSGKHQFYKTDEVIDLPEATQFEIVDSKLINQNLKLRKK*

Sequence 1627

Contig_0629_pos_890_252,
is similar to (with p-value 1.0e-44)
>sp:sp|P16440|RISA_BACSU RIBOFLAVIN SYNTHASE ALPHA CHAIN (EC
2.5.1.9). >pir:pir|S45544|A35711 riboflavin synthase (EC 2.
5.1.9) alpha chain - *Bacillus subtilis* >gp:gp|L09228|BACDIA_11
5 *Bacillus subtilis* spoVA to serA region. NID: g410114. >gp
:gp|X51510|BSRIB_3 *B. subtilis* riboflavin biosynthesis operon
ribG, ribB, ribA, ribH, and ribT genes. NID: g40083. >gp:gp
|Z99116|BSUB0013_39 *Bacillus subtilis* complete genome (secti
on 13 of 21): from 2395261 to 2613730. NID: g2634723.
10 atgtctatgtttacaggtatcattgaagaaataggtactgtacaacaagttcgctctgaa
caatcagtaagaacgcttgaaattaaagcacaaaaacatttttagttgatatgcatattggt
gattcaataaagtgttaacggtgcatgtttaactgtgatagatttcactgactcaagtttt
tcagttcaagtcataaagggaactgaaaacaaaacatatcttggaagtgttcaacgtaat
15 acagaagttaatctcgaaagacccatgagtggaagtgggagatttggtggacatttcgtg
ttaggtcatgttgatgagcttggaacaattttctaaaatcaatgaaactgctaactcaaaa
attatttctattaaaacaactaaaaatattttgaaatcaaatggtaaagcaaggttctata
actgtagacggaggttagtcttactgtattttgatttacctgattatacttttgatatacat
cttataaccagaaacacgtcgatctactattctttcatctaaaaaagtggcgacaaaagt
20 cacttgaggtctgacgtactattcaaatatgttgaaaacatcatgaatcaaatcaatcg
cagttaacagaagaaaagcttagagcatttggttttag

Sequence 1628
MSMFTGIIIEIGTVQVRSEQSVRTLEIKAQNILVDMHIGDSISVNGACLTVIDFTDSSF
25 SVQVIKGTENKTYLGSVQRNTEVNLERAMSGSGRFGGHFVLGHVDELGTISKINETANSK
IISIKTTKNILNQMVKQGSITVDGVSLTVFDLHDYTFDIHLIPETRRSTILSSKKVGDKV
HLESDVLFKYVENIMNQNSQLTEEKLRAFGF*

Sequence 1629
30 Contig_0630_pos_4521_0,
putative peptide of unknown function
gtgtgcatatttagtatttaagcacagaaaaagaacaattagcaggattgaaattttctatc
agtttaaaagtgattgagcgtctacttttagcactcattctaccacttatcattttaatg
attggcttgtttagctttaacttatgctgatagtttcatcctattacaaacttcagat
35 ttatcagtatcattattaactatattaattggtcatattttaatggcttttgtagtgag
tttggtttccgttcttacttacaaaatattcttgaacaagaatgaacacattttttgcg
agtattgtcgttggtcttatttattcagtat

Sequence 1630
40 VCILVFKHRKEQLAGLKFSISLKVIERLLALLILPLIILMIGLFSFNTYADSFILLQTS
LSVSLTLILGHILMAFVVEFGFRSYLQNILETMRNTFFASIVVGLIYSVF

Sequence 1631
Contig_0630_pos_2706_1588,
45 is similar to (with p-value 2.0e-83)
>sp:sp|P54955|YXEP_BACSU HYPOTHETICAL 41.6 KD PROTEIN IN IDH
-DEOR INTERGENIC REGION. >gp:gp|Z99124|BSUB0021_51 *Bacillus*
subtilis complete genome (section 21 of 21): from 3999281 to
4214814. NID: g2636442. >gp:gp|D45912|D45912_19 *Bacillus su*
50 *btilis* genome sequence between the iol and hut operon, parti
al and complete cds. NID: g1408482.
atgacaaattattctacttatgtagattggagaagaacgtttcatcaatattcctgaactt
tcagatgaagaatatgaaactacagaaaagttacgaaaaatactcaaaagttatggtata
cgtatactggaggtacctttaaaaacaggttttagtagcagaaattgggcaaggagaggaa
55 atgatagcagtaagaacagatattgatgctttgcctatagaagaacaagtgaagcatgaa
tttacaatcaaagtatcaaggtgcaatgcatgcatgtggtcatgatattcatatggcaagt
atattagctactggtattcaactaaaagagattgaagatgaattaaatggacgcgttcga
ttaatatattcaacctgctgaagaattaggacatggtgcatttgaaatcataaatactgga
gtacttaaaggagctaaagcagtacttggttttcataattatcccactttaaaagtgtg

gaatttgctattaaatcggtgcaattacctctgctgctgatcgtttgagtttaagtgt
 aaaggtaaaggtgctgcatgctgcaaacctgagcaaggaaatgatccagtcacgctgta
 ggacaacttatcaatagttttacaaactattgtgagtcgaaatttatcagcttttgatagt
 gcagttgtaacaatcggtgaaatttcttggtggaacacatggaatgttatagctgacaaa
 5 gcttatatacagggcactgttcggttcattcgatgaggatatacgtcatttatattgaaaat
 aggatgaaaaatattgctgatggtttaagtcgtgttttaagtgtggatattgatttaact
 tattcaagactacctgggtgacagtagtaaatgatgcacatctaacacaagaagcaatcgag
 gtcgctaaaaatgttggctatcatgtatcaatgctcgatgaaccggttactattggagag
 gatttttcaggttatacagaagaataccccgggtgttttcgcatttatgggtccgacagt
 10 aaatatgatttacatcatcctaaatatcatccagatgagcgtattttggaaaaagttcct
 caatatttcggttcagctcgttcaacggtttattgcataa

Sequence 1632

MTNYSTYVDWRRTFHQYPELSDEEYETTEKLRKILKSYGIRILEVPLKTGLVAEIGQGEE
 15 MIAVRTDIDALPIEEQVKHEFTSKYQGAMHACGHDHMASILATGIQLKEIEDELNGRVR
 LIFQPAEELGHGAFEIINTGVLKGAKAVLGFHNYPTLKVGEFAIKSGAITSVDRFEFNV
 KGKGAHAAPQEGNDPVIIVGQLINSLQTIIVSRNLAFDSAVVTIGEISCGNTWNVIADK
 AYIQGTIVRSFEDIRHYIENRMKNIAADGLSRVFNVDIDLTYSRLPGAVVND AHLTQEAIE
 VAKNVGYHVSMLDEPVTIGEDFSGYTEEYPGVFAFIGSDSKYDLHHPKYHPDERILEKVP
 20 QYFVQLVQRLLT*

Sequence 1633

Contig_0630_pos_1491_610,
 putative peptide of unknown function
 25 atgtcagctcaagatccgcgcaataaattttaaactgataattatgaaaaacaagaacaa
 agagttccaggtatacaagctaaaatgtcaccacaaccagattgtgggaagattcttat
 catggccaccatcgattagatggctttaaaatactagtgactgggtggcgattcagcaatt
 ggacgtgctggcagcaattgcttatgctaaagaaggtgcagatgtagcgattaattattta
 ccaagtgaacaacaagatgccgatgatgtaaaacagattattgaaaatgttgggcaaaaa
 30 gctatcttaattcctgggtgatattagagatgaacaattcaactatgacatggttgaaaag
 gcttatcagcaattaggtgggttagacaatgtaacgttggttgctggtcatcaactttat
 caagtgaattatccggagtttaaactcaagattttaccgaaacgtttgaaacgaatgtc
 tatccggtattttggacagtcacaaaagcgcttgagtatttacaaccaggaggctcgatt
 acaacaacatcttcagttcaaggttataatcctagtccaattcttcattgattatgctgca
 35 acgaaagctgcaattatatctttaacaaagagtttttcagccgaacttggccctaaaggt
 attcgtgttaactgtgttgacacctggaccgttttggtcaccacttcaaattgtcgttgga
 caaccacaaagcgctatacctacttttggacaaaacacaccgttaggacgtgcccggccag
 ccagttgaatgtgctgggacatatgtgttattagcctctgatgatgcaagttatattacc
 ggtcaagtatatgggtgactgggtgggactcaaatagattaa
 40

Sequence 1634

MSAQDPRNFKFTDNYEKQEQEVPGIQAKMSPQPDGSDSYHGHHRLDGFKILVTGGDSAI
 GRAAAIAYAKEGADVAINYLPSEQDADDDVKQIENVGQKAILIPGDIRDEQFNVDMEK
 AYQQLGGLDNVTLVAGHQLYQDELSEFKTQDFTETFTNVYPVFWTVQKALEYLQPGGSI
 45 TTTSSVQGYNPSPILHDYAATKAAIISLTKSFAELGPKGIRVNCVAPGPFWSPLQIVGG
 QPQSAIPTFGQNTPLGRAGQPEVCAGTYVLLASDDASYITGQVYGVTTGGTQID*

Sequence 1635

Contig_0630_pos_0_323,
 is similar to (with p-value 7.0e-29)
 >gp:gp|Z99122|BSUB0019_105 Bacillus subtilis complete genome
 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 >gp:gp|Z93767|BSZ93767_8 B.subtilis DNA; 15.2 kb fragment,
 from ywqN gene to ywrO gene. NID: g1929325.
 55 gtgaaaaaaataaaagctgaaacactttcagatatgcaaaattataaattacttagtgga
 agtattatacctaggccgatagcatttgtaactactcaaaattttaaaggggatatcaac
 gcagccccgttttagtttttttaagttagttaatcatacaccacctatgattgcaattgct
 gttcaacgtacaaagggaatagaaaagacacctcaataaatatagaacaatcaggtgag
 tttgtagtgcattactgatgaggctattgttaatgatgtgaatgaaactgctgccccg

ttagaatatggtggttaatgaGTG

Sequence 1636

VKKIKAETLSDMQNYKLLSGSIIIPRIAFVTTQNLKGDINAAPFSFFNVVNHTPPMIAIA
5 VQRTKGNRKDTSINIEQSGEFVVHITDEAIVNDVNETAAPLEYGVNEX

Sequence 1637

Contig_0633_pos_1066_1476,

putative peptide of unknown function

10 atgcacaaatataataatcaaaaatataaagagggaatacacagtgcctgtgcatgttcaa
cacatgattgtctcaattcatgataagtgaggaaatttgaaagacatttgaataaaatgcga
aagatatatagagataaacttgattatattttaaaacgattaaagccctacaataactcaa
ataaagattgaaggcgactaactggaatgcattttacaataactgttaataatggattg
tcaatgaaacaatgtttaaaaaatgcgaaaaaaaataattttaaaattaaaaccttatcat
15 tacgaaaattattctaaagttttccaaaattttatttttaggatttggggggataaaaaaa
gaagaattagaagatcatgttaatgcattaattcattcactcgttatataa

Sequence 1638

MHKYNNQKYKEGNTVPVHVQHMIQAQFMISGKFERHLNKMRIYRDKLDYILKRLKPYNTQ
20 IKIEGALTGMHFTITVNNGLSMKQCLKNAKNNLKLKPYHYENYSKVYPKFI LGFGGIKK
EELEDHVNALIHSLVI*

Sequence 1639

Contig_0634_pos_695_1129,

25 is similar to (with p-value 5.0e-21)

>sp:sp|P30267|YKAA_BACFI HYPOTHETICAL 50.9 KD PROTEIN IN KAT
A 3'REGION (ORF A). >pir:pir|S27491|S27491 hypothetical prot
ein A - Bacillus firmus >gp:gp|L02548|BACKATA2_1 B.firmus OR
F A and ORF B, complete cds. NID: gl43118.

30 gtgatgaaaaatatttgggtgaatttcaaagaaagtctgattatgactatgaatatctta
cccaccatattatcaataggtttaatttgccttgctactgcagaatatacagtgaatttc
gattatttagcatatgttttttatccattaaacttgataacttcaaataaccagatttcctt
ttaactgcaaaaggcgagctattggtataacagaaatgttttgccttcattaattgta
gtcgaagcaccattaatcactaaatttataattgctgttacttctgtttctacaattata
35 ttcttttcagctagtgtgcctagtattctctctactgatatacccatccgcataagagat
ttagtggttatatgggttgagagaactgtattgagtttaattatagtaaacacctatcgca
tatatttttttataa

Sequence 1640

40 VMKNIWLNFKESLIMTMNLTILSIGLICLLLAETVIFDYLAYVFYPLTWILQIPDSF
LTAKGAAIGITEMFLPSLIVVEAPLITKFIIAVTSVSTIIFFSASVPSILSTDIPIRID
LVVIWFERTVLSLIIVTPIAYIFL*

Sequence 1641

45 Contig_0634_pos_1506_1973,

is similar to (with p-value 1.0e-76)

>gp:gp|L23109|STASINA_1 Staphylococcus aureus recombinase (s
in) gene, complete cds. NID: g495088.

50 gtgagaatgggagatcggtttgtcgtggaatcaattgatcggttaggtcgtaattatgat
gaagtcattaataaccgttaattatcttaaagagaaagaagtacaattgatgattaccagt
ctaccaatgatgaatgaagtcatttggaatccattattagataaaattatgaaagattta
attatacagatatagcaatggttcagaacaagaaagaatgaaagtaaacgtcgacaa
gctcaagggttcaagttgcgaaagagaaaggcgtatataaaggacgacctttattatat
tctcccaatgcgaaagatcctcaaaaacgtgttatctatcatcggtgtgtagaaatgta
55 gaggaaggcaaaagcgattagtaaaattgcgaaagaagtgaattatacagacaaactgtt
tatagaattaaacacgataacgatgatttgggttctgttgcaaagtaa

Sequence 1642

VRMGDRFVVESIDRLGRNYDEVINTVNYLKEKEVQLIMITSLPMMNEVIGNPLLDKFMKDL

IIQILAMVSEQERNESKRRQAQGIQVAKEKGVYKGRPLLYSPNAKDPQKRVIYHRVVEML
EEGKAISKIAKEVNITRQTVYRIKHDNDLGSVAK*

Sequence 1643

5 Contig_0634_pos_1281_862,
putative peptide of unknown function
atgattagggtgtaccctaaatatataattgagtttaaaataaaaaacagtaaattaaga
tggaatttcatatccttgctttccttatcactataaatatttgtaaaaaaggtgatttt
10 tttatatacgataattttattattgaattataaaaaaatatcgcataggtgttactata
attaaactcaatacagttctctcaaaccatataaccactaaatctcttatgcggatgggt
atatcagtagagagaatactaggcacactagctgaaaagaatataattgtagaaacagaa
gtaacagcaattataaatttagtgattaatggtgcttcgactacaattaatgaaggcaaa
aacatttctgttataccaatagctgcgccttttcagttaaaaaggaatctggtattga

Sequence 1644

15 MIRVYPKYIIEFKIKNSKLRWKFHILAFILITINICKKGDFFIYDILLNKKIYAIGVTI
IKLNTVLSNHITTKSLMRMGISVERILGTLAEKNIIIVETEVTAINLVINGASTTINEGK
NISVIPIAAPFAVKKESGI*

Sequence 1645

20 Contig_0635_pos_1358_1813,
putative peptide of unknown function
atggtaagaagaatagaagatcacatctcatttttagaaaaatttattaatgatgtaaat
25 acattaacggcaaggttacttaagacttgcaaaactgagtaggcatatcagctgagcaa
tctcatgtgttaaatatgcttagtatagaggcgttaactgtggggcaaaattacagagaaa
caaggtgtaataaaagctgctgttagtcgaagagtcaaaaagttgctcaatgctgaatta
gttaaattagaaaaacctgattccaatactgaccaacgtcttaaaataattaaattatct
aataaaggaaaaaaatatattaaagagagaaaaagcgattatgagccatattgctagtgat
30 atgacgagtgactttgacagtaaggaaattgaaaaagtttagacaggttttagaaattatc
gactatcgatatacaatcttatacttctaaactttga

Sequence 1646

35 MVRRIEDHISFLEKFINDVNTLTAKLLKDLQTEYGISAEQSHVLNMLSIEALTVGQITEK
QGVNKAASRRVKKLLNAELVKLEKPDSENTDQRLKIIKLSNKGKKYIKERKAIMSHIASD
MTSDFDSKEIEKVRQVLEIIDYRIQSYTSKL*

Sequence 1647

40 Contig_0635_pos_2075_3481,
is similar to (with p-value 3.0e-35)
>gp:gp|AF051917|AF051917_2 Staphylococcus aureus plasmid pSK
41, complete sequence. NID: g3676412.
atgagtagaggtgataaaatgagtcgaatgcccaaattgtggtcatcaagtgaagatgat
acatcgcaatgtccaaactgtgggcaactattaactaagaagaaaaaagaaagattaaa
45 gaccaatcatctcaatcgagtaatgagaattctaccaatatacgtcttcgtaaaattgtg
ccgataggattatgtgtatttatcttaatacttattatcgtgttatttttccttttaaga
aattataattcgccatgacaaagctaagatattagttaatgctgtagataataatgat
tcacaaaaagttgctacattattgagtactaaaaataaaaaagtagacgatgttgaaagcg
caacaatatattaattatgtaaaaaaagaagtaggtattaagaagtatatcagagatc
50 aataaactgtagataaaattgaataaaaagtaattcaagcgtggcatcttatatacaaacg
aaaagtggaacagatgtacttaagataagtaaaaatggtacaaagtatttaatttttgat
aatatgagtttcacagctccgactaaaaagccaattattaaacctaaagtagaaactaaa
tatgaatttagaacaagtggttaaaaaagaaaactgtcattgctgaagcaaaataaaaaataca
cctttgggtgaattttattcctggtacatatcatttaccagctaagaaaattacagaaaaac
55 ggtacattcaatgggcatttaattttgacttttagagaaagccactctgaaaccgtagat
gtagctgaagattatgatcaatcattttatcaatatcaaatttaaggggtgcgaataaatta
agtataaatcagaaaaagttcaaatcaatgaccgtacattcacttattctcattctaaa
gaatttggtccttatccaaaaacaaaagatataacgatttctgcaactggtaaggcaaaa
ggtaagacgtttagttctgagacgaaaacaattagtgacagacgatttgaaagataatacag

aaagttacattggaatttgatagtgataaaataaatagctatggtgagaagaaagaaaaa
gaagaaaatagtttgaaaaataaattaactgaattttttactgggtatgcaacggctatg
aattcagcatttaatatgaatgattttaactttatatcgagttattttaaaaagaattcg
tctatatacacatcaatgaaaagtaatttccaaaatcgaaacgaacgtgactatgatattc
5 ccgcaagtgttaagtgttcacgaaacggacatactgtaagaacaactattcaacatattc
gatcatattggtaattatataaataaagattatgaattagaaatagataatgatgatagt
aatatgcagttggttaaagaattataa

Sequence 1648

10 MSRGDKMSQCPNCGHQVKDDTSQCPNCGQLLTKKKKKRIKDQSSQSSNENSTNIRLRKIV
PIGISVFILILIIVLFFLLRNYNSPNAQAKILVNAVNDNDSQKVATLLSTKNKKVDDVEA
QQYINYVVKKEVGIIKYIRDINNTVDKLNKSNSSVASIQTSGQDVLKISKNGTKYLIFD
NMSFTAPTKKPIIKPKVETKYEFRSTGKKKTVIAEANKNTPLGEFIPGTYHLPAAKITEN
GTFNHGLNFDNFRESHSETVDVAEDYDQSFINKFKGANKLSDKSEKVQINDRTFTYSHSK
15 EFGPYPKTKDITISATGKAKGKTFSETKTISADDLKDNTKVTLFSDKINSYVEKKEK
EENSLKKNLTEFFTGYATAMNSAFNMNDFNFISYFKKNSSIYTSMKSNFQNRNTNVTMIS
PQVLSVHRNGHTVTRTTIQHIDHIGNYINKDYELEIDNDDSNMQLVKEL*

Sequence 1649

20 Contig_0635_pos_5479_6126,
putative peptide of unknown function
atgaaaaaatgatattaatcaatgtgattactatcattgtcctagttggtattggtgtg
ttaggcttttggtctgtgcataacacacaagttatgtgacaactgacaatgcaaaagtt
gatggagatcaataaaaaatctcaagtcctgcacatctggacaaattaaatctcttaattgtt
25 aagcaaggagacaaaacttgataaaggtgataaagtagcagaagtttagcacaaggccaa
gatgggaatcaaaagatatgaacatcaaaatgccacaaaaggtactattggtataaaca
gatggtatcgaaggttctatgactcaagcaggggaacccaattgcatatgcatataattta
gatgatctatatataactgctaattgtagatgaaaaagatatcttctgacgtggaaaaaggc
aacgacgttgatgtagatatcgacggtcaaaaagcatcaatcaaaggtgaaggtgaagaa
30 gtaggccaaagcgactgcagctagcttttcattgatgccatcatcaaatagcgacggtaac
tatacgaaagtttctcaggtagtaccgtaaaaaatctcttttagattctaatccatctaaa
aatgttgctccaggtatgaacgctgaagttaaaaattcataaaaattaa

Sequence 1650

35 MKKMILINVITIIIVLVVIGVLGFWFHNNTSYVTTDNAKVDGDQIKISSPASGQIKSLNV
KQGDGLDKGDKVAEVLQAQQDQGSQKDMNIKMPQKGTIVKTDGIEGSMTQAGNPIAYAYNL
DDLITANVDEKDISDVEKGNVDVDIDGQKASIKGKVEEVGQATAASFSLMPSSNSDGN
YTKVSQVVPVKISLDSNPSKNVPGMNAEVKIHKN*

Sequence 1651

40 Contig_0635_pos_6139_7053,
is similar to (with p-value 9.0e-47)
>sp:sp|P54585|YHCA_BACSU HYPOTHETICAL 58.3 KD PROTEIN IN GLP
D-CSPB INTERGENIC REGION. >gp:gp|Z99108|BSUB0005_169 Bacillu
45 s subtilis complete genome (section 5 of 21): from 802821 to
1011250. NID: g2633055.
atgacagcgaccttcattattatatattgtagtagcgctcactcattggttttatt
aatttctttttaattaagcgtaaaagaaaaataaagacaaaagagtggaacaacgttcg
acaatagattctaagagagaaagcaatcaatctaaatttaaagcaagcgatttagaacia
50 acaactaagtcaataactgacccaacgcaatcaaacgatattgaagatgaaaaacgaaaa
aatcactttgactcagaaatagataatgcatctcaatctatcaatacagatagtaaagag
gatagaaacgcgttaagtcataagaaccaagaggaagatgacgcatcgaacgatgtgttg
aaccctatcgatccaaattctactgaaggtagagttaatgaaagaattaaaaatcaagag
tctaactttatttttggtaaaggcataactagaggtaaaatttttagcggcaatgttattt
55 ggtatgtttatcgcgattctaaaccaaactctattaaatgtggcattgcctaaaataaat
acagagtttaataattttctgcctcaactgggtcaatgggttaatgactggttttatgtagtg
aatggtatattaatacctattagtgctttttatttaataaatattcttatagaaaatta
tttattataggttagcactatttacattaggttccttagtttggtgcaatctcatttaatt
tcccaattatgatgagtgacgtgtattacaagccataggcgaggtatattgatgccg

ttaggttctaacgttattgttaccattttccacctgaaaaacgcggtgtggcaatggg
acaatgggtattgcaatgatattagcacctgcaatcggtccaacactttcaggttacata
gtgttcggttactag

5 Sequence 1652

MTATFIIIIYIVVALILIGFINFFLIKRKRKNKDKRVEQRSTIDSKRESNQSKFKASDLEQ
TTKNTDPTQSNIDIEKRNHFDSEIDNASQSINTDSKEDRNALSHKNQEEDDASNDVL
NPIDPNSTEGRVNERIKNQESNFIFGKGITRGKILAAMLFGMFIAILNQTLNVALPKIN
TEFNISASTGQWLMTGFMLVNGILIPISAFLEFNKYSYRKLFIIIGLALFTLGLSLVCAISFN
10 FPIMMSGRVLQAIGAGILMPLGSNVIVTIFPPEKRGVAMGTMIAMILAPAIGPTLSGYI
VLRV*

Sequence 1653

Contig_0635_pos_7086_7757,

15 is similar to (with p-value 3.0e-49)

>gp:gp|AF044668|AF044668_5 Salmonella typhimurium (g30k) gen
e, partial cds; and 50S ribosomal protein L32 (rpmF), PlsX (
plsX), 3-oxoacyl-acyl carrier protein synthase III (fabH), m
alonyl CoA-acyl carrier protein transacylase (fabD), and 3-o
20 xoacyl-acyl carrier protein reductase (fabG) genes, complete
cds. NID: g3282798.

atgggacatagcttaggagaatattcaagcttagttgctagtggtgtattatctttgaa
gatgcggttagaattgtgcgtaaacgtggccaacttatggctcgagcggttcctaacggt
gttgagggtatggcagcagattaggtttgattatgatgatgttgataagatatgtcaa
25 acgttatctacaaaagaacagttaattgaacctgctaataactcaccaggtcaaact
gtggtgtctggacataaatctttaattgatgaattagtagaaaaggcacaagaacttgg
gttaaacgcgttcttccattagctgtttccggtcctttcattcttcaatgatgaaagt
attgaagaggattttgctaatttcattaatcaatttgaatggcataatgctaattatcca
gttggttcagaatgttaatgcaaaggagaaaccgatgctgaagtaattaaacgcaatatg
30 gttaaacaattatattcacctgttcaatttattcaatcaacggagtgggttgattaatcaa
gggtgtcgatcactttattgaaattggaccgggaaaagtattatctggacttatcaaaaa
ataaatcgagatgtaaaaatcacttcaattcaaacactcgaagatgtgaaaggatggaat
aatcatgaataa

35 Sequence 1654

MGHSLGEYSSIVASGVLSFEDAVRIVRKRQQLMARAFPNGVGGMAAVLGLDYDDVDKICQ
TLSTKEQLIEPANINSPGQIVVSGHKSILDELVEKGKELGAKRVLPLAVSGPFHSSMMKV
IEEDFANFINQFEWHNANYPVVQNVNAKGETDAEVIKRNMMVKQLYSPVQFIQSTEWLINQ
40 GVDHFIEIGPGKVLSGLIKINRDVKITSIQTLEDVKGWNNHE*

Sequence 1655

Contig_0635_pos_7969_8484,

is similar to (with p-value 2.0e-65)

>sp:sp|P51831|FABG_BACSU 3-OXOACYL-[ACYL-CARRIER PROTEIN] RE
45 DUCTASE (EC 1.1.1.100) (3-KETOACYL- ACYL CARRIER PROTEIN RED
UCTASE). >gp:gp|U59433|BSU59433_3 Bacillus subtilis PlsX (pl
sX), malonyl-CoA:Acyl carrier protein transacylase (fabD) an
d 3-ketoacyl-acyl carrier protein reductase (fabG) genes, co
mplete cds, and acyl carrier protein (acpP) gene, partial cd
50 s. NID: g1502418.

gtggtaagtcagtttgggttctgtagatgtattggttaacaatgcagggataactaaagac
aacttacttatgcgtatgaaagaacaagaatgggatgacgtgattgatacgaacttaaaa
gggtgtgtttaactgtattcaaaaagtaacgccacaaatgttgcgtaacgtagtgtgca
atcattaatctaactagtattgttggtgcaatgggtaatcctggacaagcaaaactatgtt
55 gcaacaaaagcaggtgtcattggattaacaaaaactgcagcacgagaactagcatcacga
gggtattacagtgaacactgtagcacctggtttcatcgtttcagacatgacaaatgcttta
agtgtatgatttgaaggatcaaatgttagagcaaatcctttaaacgttttggagaagat
acagatatagctaatactgttgcttcttagcttctgataaagctaaatatattacaggc
caaacaattcatgttaacggtggaatgttatgttaa

Sequence 1656

VVSQFGSV DVLVNNAGITKDNLLMRMKEQEWDVIDTNLKGVFNCIQKVT PQMLRQRSGA
IINLTSIVGAMGNPGQANYVATKAGVIGLT KTAARELASRGITVNTVAPGFIVSDMTNAL
5 SDDLKQMLEQIPLKRFGEDTDIANTVAFLASDKAKYITGQTIHVNGGMYM*

Sequence 1657

Contig_0635_pos_9060_9797,

is similar to (with p-value 2.0e-63)

10 >sp:sp|P51833|RNC_BACSU RIBONUCLEASE III (EC 3.1.26.3) (RNASE III). >gp:gp|D64116|D64116_3 Bacillus subtilis genes for ORF1, ORF2, ORF3, ORF4 and Srb, partial and complete cds. NID : g1389548.

gtggctaaccagaaaaagaaagaatggtacatgattttcaacaaaaatttactgataaa
15 atgaagtcgtaggattacgttttaaaaaatattgatttatatcaacaggcattctctcat
tcaagttttattaatgactttaatatgaatcgtagaacacacgaacgcttagaattt
ttaggtgatgcggtattagaattgacggtttcacgctatctttttgacagacatcctcat
ttaccagaaggaatttgacaagatgcgcgcaacaattggttgtaaccttcacttggtg
atatttgcgaataagattaaattaaacgaactgattttattaggtaaagggtgaagagaag
20 acaggaggcagaacaagaccttcccttatttcagatgcatttgaagcttttgtaggtgca
ctgtatttagatcaaggtttagattcagtatggacatttgctgaaaaagtcattcttccg
tatgtagaagatgacgagcttgttggtgctgtagactttaaaacacagttccaagaatat
gtgcatagccaaaataaaggagatgtgacataccaattaattaagaagagggtcccgc
catcatagactatttacatcggaagtgattttagaaaataaagcagttgcagagggtaaa
25 ggaaagacaaagaaagaatccgaacaaaaggcagcagaacaagcgtataaactaatgaaa
aataaaaaatcattataa

Sequence 1658

30 VANQKKKEMVHDFQOKFTDKMKSLGLRFRKNIDLYQQAFSHSSFINDFNMRLEHNERLEF
LGDAVLELTVSRYLFDHRPHLPEGNLT KMRATIVCEPSLVIFANKIKLNELILLKGEEK
TGGRTPLSLISDAFEAFVYGALYLDQGLDSVWTF AEKVIFPYVEDDELGVVDVFKTQFQEY
VHSQNKGDVTVYQLIKEEGPAHHRLFTSEVILENKAVAEGKGKTKKESEQKAAEQAYKLMK
NKKSL*

35 Sequence 1659

Contig_0635_pos_10530_0,

is similar to (with p-value 0.0e+00)

40 >sp:sp|P51834|YLQA_BACSU HYPOTHETICAL 135.4 KD PROTEIN IN RNC-SRB INTERGENIC REGION (ORF4). >gp:gp|D64116|D64116_4 Bacillus subtilis genes for ORF1, ORF2, ORF3, ORF4 and Srb, partial and complete cds. NID: g1389548.

gtggaaccggttaaaagaagaggcgccattgctaaagaatataagcaattatctaaagag
atggagcaaagtgtatgtcatcggttacagtatctgacattgatcattatactgaagataat
cagcgatttagatgagcggttaaatcacctaaagagtcaacaggctgagaaagaaggtcaa
45 caagctcaaatcaatcagttactacaaaaatataaaggtaaacgtcaacaaaacgattat
gacattgaaaagttaaattatgaattagttaaagcaactgagaattatgagcaattatca
ggtaagctaaatgtattagaagaacgaaagaaaaccaatcagaacaaaatgcaagatat
gaggaagaattagataatttagaatcacaaattgattctattaaaaatgaaaaagcccaa
aatgaaaaattattagctgagttaaaaataagcaaaagcaattaaacaagggaagttcaa
50 gaattagagtcacttctttatatatccgatgaacaacacgacgaaaaactagaagaatt
aaaaatagttattatacattgatgtcagaacaatcagatgttaataatgatataagattt
ttagaacatacaatcaatgaaaatgaagcaaaaaaatcacgatttagattcgcggttagta
gaggctttcaatcaactaaagacattcaacaaaatattactcaaacacaaaaggaatac
caaagttctaagaaatctatggaaaaagtagaacaataattcaacaattagaacaacag
55 ttgacagattctaaaagacttctatctgaatatgaaaataaactatatcaagcctatcgt
tataatgaaaagttaaaatcaagaattgatagtttagctactcaagaggaagattacacg
tatttctttaatggtgtaaagcacattttgaaagcaaaagataaagaattaagaggaatt
catggtgcggttgacagaagtgattaacgttcttcagaaatgacacaagcgattgaaacc
gccttaggtgcatcggttacagcacgattattgttgataatgaaaagacggtcgccaagca

atccagtgacttgaagcaaagaggttttaggtcgtgctacttttttaccattaaacgtgatt
 caaccaagacatgtagctgctgacattaaagatgtagctcgtggttcacaagggttcatt
 aatattgcatctgatgccataaatgtatctgctaaatatcaaaacatcattgaaaattta
 ttaggtaataccatcattgtagaaaatttaaaacatgcaaatgaacttgcacgtgccatt
 5 cgatatcggacaagaatagtaacttttagaagggtgatggttgtaaatcctgggtggtccatg
 acaggaggaggagcacgtaaaacaaaaagtatatgtcacaaaaagatgaattatcaaca
 atgcgaaatcaacttgaagattatcaacgacaaacagcagaatttgaacgtcagttttaa
 gaacaaaaaacacaagctgaacaattaagtgaacaatatcttagtgcaagtcagcagtac
 aacaattttaaagaacaagtagatcatcacgaattagaactggatagactaaaaacacaa
 10 gaagcacatcttaaaaatgaacatgaagagtttgaatttgaaaaaatgatggatatcaa
 agtgataaaaagtaaaagaaacattaaaagaaaaacaaatcatttaattgagatacaacaa
 caatttgagcaactagaaaagttagattgaaagatatatacacaattatcaaaagaaggaaaa
 gcttcgacacatcaaacacacaaactacatcaaaaacaatctgatttagctgtgtt
 aaagagcgaattaaatcgcaaaagcaagtttatgaacgttttagataaacaacttagcgat
 15 tcagaacgtcaaaaaattgaagtaaatgaaaaaatcaaatgtttaattcagatgaatg
 atgggtaaagatgcttttgaaaagttgaaagagcaaatcagcaacaagaaaatgtaaga
 caaaattttaaatacaacaacttagtgagattaaacagcaacgtaaaagatcttaatgagaaa
 atcgaaataaatgaaagtcagcttcaaaaatgtcatcaagatatactttctatagaaaat
 catttgaagatatataagcaaaacaatcaaaagctagatgtattaatcaaccatgcaata
 20 gatcattttaaattgacacgtatcaactcacagtagaacgtgcaagaatggaatatgatct
 gatgaaactattgacaatttgcgtaaaaaagtaaaattaacgaagatgacaatcgatgaa
 ttaggtcctgtaaatttaaatgcaatagaacaatttgaagagttgaatgaacgatacaca
 tttttaaatgagcaacgaacagatttaagagaaagcaaaagaaaccttagaacaacatt
 catgaaatggataaagaagttgaaggtcgttttaagacaacatttcatgcggttcaagat
 25 cattttacgacagtggtttaagcaattatttgggtgggacaagcagaacttcgtttaact
 gaagatgactatttctgctgctggcgttgatatcatcgtaaccgccaggaaaaaaatta
 caacatctttcatta

Sequence 1660
 30 VEPLKEEAAIAKEYKQLSKEMEQSDVIVTVSDIDHYTEDNQRLDERLNHLKSQQAEEKEGQ
 QAQINQLLQKYKGKQNDYDIEKLNLYELVKATENYQLSGKLVLEERKKNQSETNARY
 EEELDNLESQIDS IKNEKAQNEKLLAELKNKQKQLNKEVQELESLLYISDEQHDEKLEEI
 KNSYYTFLMSEQSDVNNDIRFLEHTINENEAKKSRDLSRLVEAFNQLKDIQQNITQTQKEY
 QSSKKSMEKVEQNIQQLEQQLTDSKRLLESEYENKLYQAYRYNEKLKSRIDSLATQEEDYT
 35 YFFNGVKHILKAKDKELRGIHGAVA EVINVPSEMTQAIETALGASLQHVIVDNEKDGRQA
 IQYLKQRLGRATFLPLNVIQPRHVAADIKDVARGSQGFINIADAINVS AKYQNI IENL
 LGNTI IVENLKHANELARAIRYRTRIVTLEGDVVNP GGSMTGGGARKTKSILSQKDELST
 MRNQLEDYQRQTAEFERQFKEQKTAQEQLSEQYFSASQQYNNLKEQVHHHELELDRLKTO
 EAHLKNEHEEFEEFKNDDGYQSDKSKETLKEKQNLHIEIQQLKQLES DIERYTQLSKEGK
 40 ASTHTQQLLHQKQSDLAVVKERIKSQKVYERLDKQLSDSERQKIEVNEKIKLFNSDEM
 MGKDAFEKLKEQIQQENVRQNLNQQLSEIKQQRKDLNEKIEINESQLQKCHQDILSIEN
 HYQDIKAKQSKLDVLINHAIDHLNDTYQLTVERARMEYDSDETIDNLRKKVKLTKMTIDE
 LGPVNLNAIEQFEELNERYTFLNEQRTDLREKETLEQIIHEMDKEVEGRFKTTFHAVQD
 HFTTVFKQLFGGGQAE LRLTEDDYLSAGVDIIVQPPGKKLQHLSL

Sequence 1661
 Contig_0635_pos_5254_4787,
 putative peptide of unknown function
 50 atgatatgtgcatatagtgccgtgaatcgctctactttttatgatcattttcaagataaa
 tatcaattactagataagatccaaaattatcatttaacaaatatatatctttactacaa
 tctttctataacgattttcatcatatcaaaacagatcaaaaaaattatataaatttttc
 ttattgatagccaaatatattaaacgtaaaagaagcgttctacagagcaacacttgaaca
 tatcctaataaagatattgcattagattacattaaacgccactaaaacatggttatgaaaa
 gtcattgaatagatatgaacctcaataaataaataaacgtatgtttatcatttattcagtc
 55 ggtggtcaagcaggtgtattatcgattggttacgtaattggatgcatcgaatctcctcaa
 gaggtcgctcaagttcttttagctaatacaattaaattacaacgataa

Sequence 1662
 MICAYSGVNRSTFYDHFQDKYQLLDKIQNYHLNKYISLLQSFYNDFHHIKTDQKKLYKFF

LLIAKYIKRKEAFYRATLVTPYPNKDIALDYINATKTKCYEKVMNRYETSINNKRMFIIYSV
GGQAGVFIDWLRNGCIESPQEVAVQVLLANTIKLQR*

Sequence 1663

5 Contig_0636_pos_281_1405,
is similar to (with p-value 2.0e-92)
>sp:sp|P54542|YQJE_BACSU HYPOTHETICAL 39.7 KD PROTEIN IN GLN
Q-ANSR INTERGENIC REGION. >gp:gp|D84432|BACJH642_260 Bacillu
s subtilis DNA, 283 Kb region containing skin element. NID:
10 g2627063. >gp:gp|Z99116|BSUB0013_102 Bacillus subtilis compl
ete genome (section 13 of 21): from 2395261 to 2613730. NID:
g2634723.
atgattaatcaaaaacgtttattagattgttttctagaattagtgcaaattgattcggaa
acagggcacgaagaacaattcaaccttatcttaagatacgtttgaaaaaatggggctc
15 catgttattgaagatgaagcttcaaaaaataatagattaggtgctaacaatcttatttgt
acgttaaaaaagtaataataagtcatacagaatgtgccgaaaattttttacaagccacatg
gatactgtcgttccaggaaaaaacatccaacctgtagtaaaagaagatggatacgtttat
agcgatggaactacgatactcggggcggacgataaaagccggtcttgcggcaataattgaa
gcgattaacaaataaaggaatcaaatttgcacacggacagattcaaataattattacc
20 gtgggagaggaatctggattagtaggtgctaaagcaatagatactcgccttcttgatgca
gatttcggctatgctgtagatgcaagtaaagatgttggaactactgttatcggtgctcca
actcaagtaaagatttatacaactataaaaagggaacccgcccattgcaagcacacctaaa
aaaggtattagcgcaataaataattgcatcaaaaagcaatcagtcgaatgaaattgggacaa
gtcgatgcattaacaacagccaatataggtaaatttcacggaggttggtgccaactaatatt
25 atagctgatgaagtcacttttagaggcagaagcaggtcacatgatgatcaaagcattaat
aaacaagtgaacatatgaaagagactttcgaaacgacagcaaatgaattagcggtcaa
gctgaagtgttagttgaaaaaagttatccgggatttgaagttagtgaagctgacaaagta
acacaatatgctatatctagtgcaattagccctcgggtctaaaaggtgatacttgattgct
ggtggtggttcagacggcaacatcatgaatcaatatggcattccctctgtgatttttagga
30 gtaggatatgaaaacatacactacttccggaagaatagcaataaaggatatgtatatg
ctcacaagacaaataataaaaaattattgagctagtagctgaataa

Sequence 1664

MINQKRLLDLCFLVQIDSETGHEETIQPYLKDTFEKMLHVIDEASKNNRLGANNLIC
35 TLKSNISHQNVPKIYFTSHMDTVVPGKNIQPVVKEDGYVYSDGTTILGADDKAGLAAIIE
AIKQIKESNLPHGQIQIITVGEESGLVGAKAIDTRLDDADFGYAVDASKDVGTTVIGAP
TQVKIYTTIKGKTAHASTPKKGISAINIASKAISRMKLGQVDALTANIGKFHGGCATNI
IADEVTLAEARSHDDQSINKQVKHMKETFETTANELGGQAEVLVEKSYPGFEVSEADKV
TQYAISSALALGLKGDTCIAGGSDGNIMNQYGPSVILGVGYENIHTTSERIAIKDMYM
40 LTRQIIKIIELVAE*

Sequence 1665

Contig_0636_pos_1485_2891,
is similar to (with p-value 0.0e+00)
45 >sp:sp|P14062|6PGD_SALTY 6-PHOSPHOGLUCONATE DEHYDROGENASE, D
ECARBOXYLATING (EC 1.1.1.44). >pir:pir|S04397|S04397 phospho
gluconate dehydrogenase (decarboxylating) (EC 1.1.1.44) - Sa
lmonella typhimurium >gp:gp|X15651|SEGND1_1 S. enterica gnd
gene for 6-phosphogluconate dehydrogenase. NID: g47699. >gp:
50 gp|M64332|STYGND1_1 S.typhimurium (strain LT2) 6-phosphogluc
onate dehydrogenase (gnd) gene, complete cds. NID: g154099.
atgacacaacaaattggagtagtgggttttagcagtaattgggaaaaacctagcttggaa
attgaatcacgtggttatagtggttctgtttataaccgatcaagacaaaaaactgatgaa
atggttaaagaatcgctggaagagaaatttaccacatactcattagaagaatttga
55 gaatcttttagagaaacctcgtaagattttattaatggtaaaagctggacctgcaacagat
gccactatagatgggtttattacctttattagacgatgatgatattttaattgatggtggt
aatactaattaccaagatacattcgctcgaaataaagcttttagctgaaagtagtattaac
tttattggtatgggagtttctggtggagaaatcggcgcaactcacgggcccttctttaatg
ccaggtggtcaaaaaagatgcttataacaaagtcagcgatatcttggacgcaattgctgct

aaggcacaagatgggtgcttcatgtgtaacttacattggccctaattggtgcaggacattat
 gttaagatggtagacaatggtagatcgaaatgagatgcaattaattgctgaaagttat
 gcaatgatgaaagatttattagggcatgtcacataaagaaatttctcaaacttttaagaa
 5 tggatgctggagaacttgaaagtattttaatagaattacaggtgatattttcaataaa
 ttagatgatgacaatgaagcacttgtagaaaaatattagatactgcagggtcaaaaaggc
 acaggtaaatggacttcaattaacgcactagaattaggtgttcctttaacaatcattaca
 gaatctgtatttgcgagattcatctcatcaattaaagaagaacgtgttactgcttctaaa
 tctttaaaaggacctaaagcacattttgaaggcgataaaaaaacattcttagaaaaata
 10 cgtaaggcactttatatgagtaaaatagtctcatatgcacaagggttcgctcaaatgaga
 aaagccagtgaagataatgagtggaatttgaaattaggcgaattagcaatgatttggcgt
 gaaggttgattattctgtgcacaattcctacaaaaattaaagatgcctacgataataat
 gaaaacttacaaaacttattattagacccttacttcaaaaacattgttatggaatatcaa
 gatgcactacgtgaagtagtagctactagcgtgtacaatggcgtgccaacacctggtttt
 15 tcagcaagtataaattattatgatagttatcgctcagaggatttacctgcaaaacttaatt
 caagcacaacgtgattactttggcgacatacttatgaacgtaaaagaccgtgaaggtatt
 ttccatacacaatgggtagaagaataa

Sequence 1666

20 MTQQIGVVLAVMGKNLAWNIESRGYSVSVYNRSRQKTDEMVKESPGREIYPTYSLEEFV
 ESLEKPRKILLMVKAGPATDATIDGLPLDDDDILIDGGNTNYQDTIRRNKALAESSIN
 FIGMGVSGGEIGALTGPSLMPGGQKDAYNKVSDILDAIAAKAQDGASCVTYIGPNGAGHY
 VKMVHNGIEYADMQLIAESYAMMKDLLGMSHKEISQTFKEWNAGELESYLIETGDI FNK
 LDDNEALVEKILDTAGQKGTGKWTSSINALELGVPLTIITESVFARFISSIKEERVTASK
 25 SLKGPKAHFEGDKKTFLEKIRKALYMSKICSYAQGFAQMRKASEDNEWNLKLGLELAMIWR
 EGCIIIRAQFLQKIKDAYDNNENLQNLDDPYFKNIVMEYQDALREVVATSVYNGVPTPGF
 SASINYDYSRSEDLPANLIQAQRDYFGAHTYERKDREGIFHTQWVEE*

Sequence 1667

30 Contig_0636_pos_3061_4716,
 is similar to (with p-value 0.0e+00)
 >pir:pir|S44188|S44188 alpha-glucosidase (EC 3.2.1.20) - Sta
 phylococcus xylosus >gp:gp|X78853|SXMALRAG_2 S.xylosus malR
 gene and malA gene. NID: g474175.
 35 atgaaaagaaattggtggaagaagcagttgcatatcaagtatatccacgaagttttaat
 gatagtaatggagatggaataggtgatctacctggattaattgaaaaattagattatcta
 gaaaatttaggaatagatgtcatttgggttaagcccaatgtatccatcaccaaacgatgat
 aatggatgatatttagtgactacaaaggcattatgagtgaatttgggtacaatgaacgat
 40 tttgatcaattgtttatcaagcatacatcaaagagggatgaaattaattagacttagtg
 gtaatcacacatcagatgaacaccttgggttattgaaatcaaaatclagtataaacaat
 gcaaaaagagatttggtatatttgggcagatcctaaaccggatggatctgaacctataaac
 tgggaaagatctttaatggttcaacttgggagtttgacgaatcgactaagcaatactat
 ttccatttttagcaaaaagcagccagatttaaattgggaaaatccagatgtaagacaa
 gctgtgtttgaaatgatgaattgggtggtttgaaaaggtattgacggatttagagttgat
 gccattactcatattaaaaagaatttgaagcaggagatttacctgtacctgatggcaaa
 45 aaatttgcctcagcatttgatgtagatatgaatcagccaggaatacaagaatggctcaa
 gaaatgaaagataaatcggttaagtcggtagacattatgactgtaggcgaggctaattggt
 gttactcctaattgatgctgaagaatgggtaggagaagaaaatgggaaatttaatatgata
 ttccagtttgaacatcttggttatggagtactggcgatacgaattcgatgttaaatcc
 tataaacaagtcttaaatcggttgcaaaagcaactagaaaatgtaggttggaatgcttta
 50 tttatcgaaaaccatgatcaaccacgtcgtgttcaacctggggtgatgataaaaattat
 tggatgaatcagcaactagtacgctactgcctactttttacaacagggcacacctttt
 atttaccagggtcaagaaataggtatgactaattatccatttgaaagcattgaaagtttc
 aacgatgtcgagtgaaaactgaatatcaaatagtcaaaaaagaggtggagatgtcaat
 caattactagataaatataaaatggaaaaccgagacaatgcaaggactccaatgcaatgg
 55 aataattctatcaatgctggattcactactggttaagccatggtttcatgtaaacctaac
 tatacagaattaatgttaacaacaactaaatgataagttttcgatactttcttattat
 aaagcggttaattcaactaaaaaatctgatttgatttacacctacggtaagtttaatatg
 gtcgatgctgaaaataagcaggtttttgcataacacgcacatttaaaaacaatactgta
 ttaattgtagccaatctcacaaatgaagtatcagaactaaacctaccttttgaattagat

atttcatctgtagatataaaattgcataattatcacttaaataatgatataaaatttagaccat
attaaaccttatgaatcattcgtcgttgaaatataa

Sequence 1668

5 MKRNWWKEAVAYQVYPRSFNDSNGDGIGDLPGLIEKLDYLENLGIDVIWLSMPYSPSPND
NGYDISDYKIMSEFGTMNDFDQLSSIHQRMKLIIDLVVNHTSDEHPWFIESKSSKT
AKRDWYIWAOPKPDGSEPNNWESIENGSTWFEDESTKQYYFHLFSKKQPDNLNWPDPVRQ
AVFEMMNWWFEKGIDGFRVDAITHIKKNFEAGDLPVDPGKKFAPAFDVMNQPGIQEWLQ
EMKDKSLSRDITVGEANGVTPNDAAEEWVGEENGKFNMIQFEHLGLWSTGDTKFDVKS
10 YKQVLNRWQKQLENVGNALFIENHDQPRRVSTWGDDKNYWYESATSHATAYFLQQGTPF
IYQGQEIGMTNYPFESIESFNDVAVKTEYQIVKKEGGDVNQLLDKYKMNERNARTPMQW
NNSINAGFTTGKPFHVNPNYTEINVKQQLNDKFSILSYKALIQKKSDLIYTYGKFNM
VDAENKQVFAYTRTFKNNTVLIVANLTNEVSELNLPFELDISSVDIKLHNYHLNDINLDH
IKPYESFVVEI*

15

Sequence 1669

Contig_0636_pos_6783_7460,

putative peptide of unknown function

gtgggcatcttagtatcggggtcagggatagcgagtgtaacaaataataactcacgca
20 aaagaaagtcacgattcaactcctcaaaatattaaattagtggaacgtatgatacttct
caagttgattccaaaacgatgaaacaatttaagaaatagaaaaagaagataataatttc
cacataactaaacatggaaataaagtcgtttagaagacaaattacctaataccagagaat
aaaacttcaagttattcagctgatggtagtgctgaaataatacaaaaagtaattaatttc
tctgattttgttggaatatggatgggaaagatgatggaaaaatatcggtatgggataacc
25 ttttatagtggttaatacatataacggacaacacgatgggtcaaaaagtataaaaaagggact
catgtacattgtaataagatttaacggaacaaaatctgatcatagatactgggtcaaaaaa
catcctagagcttatgtagatttttataaaaagtattgctggtatcacgcaaagcttat
aatgttcttcttggaataatgactaaatgcgatgggttgatagtagtatttatagaaaa
gggtgtcaagattgctcatcatggaaaggtaaacccaacataaaaaactggcctaaaaaca
30 gcatggtatagaaattaa

Sequence 1670

VGILVSGSGIASVQTNITHAKESHSTPQNIKLVTYDTSQVDSKTMKQFKEIEKEDNNE
HITKHGKNVVEDKLPNPENKTSSYSADGSAENNTKVINFSDVGNMDGKDDGKISDGIT
35 FYSGKSYNGQHDGQKVKKGTHVHCNRENGTKSDHRYWSKKHPRAYVDFYKSDCWYHAKAY
KCSSLGKMTKCDGLNSIYRKGVKDCSSWKGKPKHKNWPKTAWYRN*

Sequence 1671

Contig_0636_pos_7464_0,

40 putative peptide of unknown function

atgaataaaatcttaaaaaatattaataacttctattattgttatcattattaccttaaca
gtttggacttttagtggtattacttatcagaaacacaagagtgagaaaatcatcaatcac
gttatagaacgtaagggtttgggataaaaaaataaaaaatgaaaaatgagtttttaatt
ataatgggatatgctgaaaaagatatgttttttaagatcaaccatatagtgagtagatgag
45 tataacgtgacaccagcaccatggacagatgataaagaatataaggtgtgggggaaaca
ga

Sequence 1672

50 MNKILKILITSIIVIIITLVWTFVITYQKHKSEKIINHVIERKGWDKKIKNEKMSFNI
IMGYAEKDIVFKDQPYSEYEYNVTPAPWTDKEYKVWGETX

Sequence 1673

Contig_0636_pos_5872_4982,

is similar to (with p-value 5.0e-20)

55 >gp:gp|X92946|LLLPK214_18 Lactococcus lactis sp. lactis plas
mid pK214, complete sequence. NID: g2467210.

gtgaaagttttggacgttatcaagcaaatacaacaggcaattgtatatattgaggatcgt
ttgttagagccttttaatttgcaagaattaagtgattacgttggtctttctccgtatcat
ttggatcaatcttttaagatgatagttggtcagtcgccagaggaatatgcgcggtgcacgt

5 aaaatgacaatagcagcaaacgatgtagttaatggagctagtcgattaatggatggttgc
 aagaaatatcggttatgcgaattctaattgatttcgcaaatgatttttagcgattttcatggt
 atctcacctattcaagctacaacaaaaaagatgaactaaaaatacagcaacgactgtat
 ataaaattatcaacgactgaaaaatgcaccctatacatagacttcaagagactgatgat
 10 atctcttagttggctattcaagatttattcctactgagcaattatcaaattccattta
 attccagacttttttagaggatctatttagtagatggttatattaaagaacttaaacgttat
 aatgatacgagcccgatgaattattttagtagtcagctgtcctctggaacaaggtttagaa
 atatttgttgggtgtccgagtgaaagttacccttcacatcttgaaagcagatttttacct
 ggtcgccattatgcattgtttaatttacaaggtgaaattgattatgctacaaacgaggct
 15 tggtattatattgagcttagcttgcaacttactttaccttatgagcgaaatagtttatat
 gttgagatttatccacttgatatttcatttaattgacccattcactaagattcaattatgg
 ttgcctattaaacaagaatctatgatttagatgaaggttatcaaaattaa

Sequence 1674

15 VKVLDVIKQIQQAIVYIEDRLLEPFNLQELSDYVGLSPYHLQDSFKMIVGQSPEEYARAR
 KMTIAANDVVGASRLMDVAKKYRYANSNDFANDFSDFHGISPIQATTKKDELKIQORLY
 IKLSTTENAPYTYRLQETDDISLVGYSRFIPTEQLSNPFNIPDFLEDLLVDGYIKELKRY
 NDTSPYELFVVSCPLEQGLEIFVGVPSERYPSHLESRLFPRHYALFNLQGEIDYATNEA
 WYYIESSLQLTLPYERNLSLYVEIYPLDISFNDFPKIQLWLPIKQEIYDLDEGYQN*

Sequence 1675

Contig_0637_pos_2774_4171,
 is similar to (with p-value 1.0e-83)
 >gp:gp|U96107|SCU96107_5 Staphylococcus carnosus N5,N10-meth
 25 ylenetetrahydromethanopterin reductase homolog, SceB precurs
 or (sceB) and putative transmembrane protein genes, complete
 cds, and putative Na⁺/H⁺ antiporter NhaC (nhaC) gene, parti
 al cds. NID: g2735503.
 atgactcaaaagtatatagatattcctacttttttagaatctatttctactatttttagttatg
 30 gttgtcgtttagtaattggttttgttttctttaatgtcccgatacaaatattattatta
 atttcttcagcttatgcagcattgattgcacatagagtggttaaaatggaaggattta
 gaagaggggattactcatcgattgagcacggcgatgccagctatctttattatttttagct
 gttggaatcattgtaggaagttggatgtattctggaacagttccagcggttaatttactat
 ggacttaaatttttaaaccaagttatttatttagtatctgcatttataatcagtgcaatg
 35 acttcaatcgctacaggaacggcttggggatcggcatctacagcaggcattgcactcata
 tcaattgctaataaattaggtgtgccagcagggtatggctgctggtgccattattgcaggg
 gcggtttttggtgataaaatgtctccattatctgatactacaaatttggcagctcttgta
 actaaagttaataatttttgcacattaaatcgatgatgtggacaacaatccctgcttca
 ataattaggattggctatatggtttattgttggattacaatataagggagacgcaaataca
 40 caacaaattcaaaatctattaaaagaattaacaacaatttataacttgaatttttgggta
 tggatcccacttattatcatagtttttagtttaattatttagaatctctacagtaccgtca
 atgcttatctctagtatcagtgcttttagttatttggacattcagtcattcaatttaatatg
 aaagatggttttaagcttcttttgatggatttaatacatacaatgctacaccagtcctcat
 atttcagataatgctaagacgttgattgagcaggggtggtatgatgagtatgactcaaatac
 45 attgtaactatattttgtggttatgcttttgcgtggtattggtgaaaaggcaggttggtta
 gacgtaatttttagagacaatagctaaaggcgtaaaagtcagttggaacactaatattaata
 actgtagttttagtagtattatgctagtagtttgcagctggagttgcttcaatagttattatt
 atggtaggcgtacttatgaaagatatgttcgaaaagatgaatgtctcaaagtcagtggtta
 tctcgtacacttgaagattcaagtacaatgggtattgccactcattccatggggcacatct
 50 ggtatataattatgcacaccaacttaattgtttcagttgatcagttcttttatatgggcaatc
 ccatgttacttatgtgcattcattgcaataatttatggctttacaggtataggaattaa
 aaaataagtagaaaataa

Sequence 1676

55 MTQKYRYPTFLESISTILVMVVVVVIGFVFFNVPIQILLISSAYAALIAHRVGLKWKDL
 EEGITHRLSTAMPAIFIILAVGIIVGSWMYSGTVPALIYYGLKFLNPSYLLVSFAFIISAM
 TSIATGTAWGSASTAGIALISIANQLGVPAGMAAGAIAGAVFGDKMSPLSDTTNLAALV
 TKVNIFAHIKSMWTTIPASIIGLAIWFIVGLQYKGDANTQQIQNLLKELTTIYNLFWV
 WIPLIIIVLCLIFRISTVPSMLISSISALVIGTFDHQFNMKDGFKASFDGFNHTMLHQSH

ISDNAKTLEQGGMMSMTQIIIVTIFCGYAFAGIVEKAGCLDVILETIAGVKSVGTLILI
TVVCSIMLVFAAGVASIVIIMVGVLKDMFEKMNVSLSVLSRTLEDSSSTMVLPLIPWGT
GIYYAHQLNVSVDQFFIWAIPCYLCAFIAYGFTGIGIKKISRK*

- 5 Sequence 1677
Contig_0637_pos_4306_5049,
is similar to (with p-value 3.0e-99)
>gp:gp|Y16431|SAU16431_6 Staphylococcus aureus dpj, alr gene
s, partial kdpC gene and 4 ORF's. NID: g3850845.
- 10 atgggccgaattggaatgaaagatatagatgaatataaagaagtgggtgatttaattaat
aaaagagatcatttagtctttgaaggggtttcacacatttgcgagtgctgatgaacct
ggaagtctctatgaatgaacaatatattttgttcaaagagatggtaatacaagttgagaaa
ccaatttatattcattgtcaaaattctgctggatcactactcatggatggtcaattttgt
aatgcaataagattaggaatctctctttatggatactatccttcagaatatgttaaagat
15 aatgtgaaagttcatttaagaccgagtgccgagtttagtatcagaaaccgttcaagtcaaa
acgcttaaaagttggggaaactgttagttatggacgtacatttattgctgatgaagaaatg
acaattgcaattttacatttgggtatgccgacggatatttaagatcgatgcaaggtgca
ttcgtcaatgttaacgggagtgcaatgtgaagtcattggacgcatttgtatggaccaaag
atagttaaaggttctctctcatgtaaaaacgggtgaaaaagtaatacttatggataatcac
20 gttgattcaccacaatcagctgaagccgtagcaataaacaaggtacaattaactacgaa
gtattatgtaatttatcaagacgtcttccaagaatatattattatgataataatgaagag
gttactaacgaattgttaaaatag

- Sequence 1678
- 25 MGRIGMKDIDEYKEVVDLINKRDHLVFEGVFTHFASADEPGSSMNEQYILFKEMVNQVEK
PIYIHCQNSAGSLMDGQFCNAIRLGISLYGYYPSEYVKDNVHVLRPSAQLVSETVQVK
TLKVGTVSYGRFTIADEEMTIAILPIGYADGYLRSMQGAFFVNVNGSQCEVIGRICMDQM
IVKVP SHVKTGEKVI LMDNHVDSPQSAEAVANKQGTINYEVL CNLSRRLPRIYYDNNEE
VTNELLK*

- 30 Sequence 1679
Contig_0637_pos_1119_769,
putative peptide of unknown function
- 35 atgttaatttatttattaagtttatttactgggtatcattggctcttttgattatttggtta
ctcaagcgtaaagagtcctcgacttattgatgtatctgggaaaacgtatctgaactatttt
atttcttatactatctattcaacagtaggcgtgatgtatgtttatgattgttccttta
atgaatataagtgaagtttagccatcttattattaattttgctgctgggtgtagtcttc
atcttattggcattgttaataatgtcatttgtgtgtacaattattgcttgcgtaaaaatat
40 atgtctggcaaaacttacactatcccactcacgatacctttataaaaataa

- Sequence 1680
MLIYLLSLFTGIIGPLIIWLLKRKESRLIDVSGKTYLNYFISYTIYSTVGVICMFMIIVPL
MNISESLAILLLILLVVVFILLALLIMSFVCTIIACVKYMSGKTYTIPLTIPFIK*

- 45 Sequence 1681
Contig_0637_pos_0_447,
is similar to (with p-value 2.0e-31)
>gp:gp|U96107|SCU96107_3 Staphylococcus carnosus N5,N10-meth
ylenetetrahydromethanopterin reductase homolog, SceB precurs
50 or (sceB) and putative transmembrane protein genes, complete
cds, and putative Na⁺/H⁺ antiporter NhaC (nhaC) gene, parti
al cds. NID: g2735503.
- 55 atgaaaaaaatcaaaacaatctcgacattggtagctggacttggatatagcatttctaggt
cacacaacacatgcagatgcggctgaaaaataacaatcaacaacaaagtacataactat
agtacaactgaagtatcattttctaattcaggaaatttatatacttctggccaatgtact
tggtatgtttatgataaaactgggtgaaaaatcgatcaacatgggggaatgcaaatagc
tgggcaactgcagctcaagcagcaggattcactgtaataatacacctgaagaaggtgca
attatgcaatcatctgaaggtgctttcggacatgttgcttctggtgaaagtgtcaataat
gatggttctattactgtatcagaaatgaactatgatgggtggccattcgctataagcaca

cgaacaatctctgccagtgaaTATAGT

Sequence 1682

5 MKKIKTISTLVAGLGIAFLGHTTHADAAENNNQQQSTYNYSTTEVSFSNSGNLYTSGQCT
WYVYDKTGGKIGSTWGNANSWATAAQAAGFTVNNTPEEGAIMQSSEGAFGHVAFVESVNN
DGSITVSEMNYDGGPFAISTRITISASEYS

Sequence 1683

Contig_0640_pos_4683_6494,
10 is similar to (with p-value 8.0e-81)
>gp:gp|D88209|D88209_1 Bacillus licheniformis DNA for Pz-pep
tidase, complete cds. NID: gl651215.
atgagtgaagggtttacctttgagagaagaagttccggtaaaagaaacttgggatttgaaa
gatttattttacaagtgatcaagcattctatcaaacattggaacaagtagtacaatgtct
15 ttagatttttaatacatacatattatcagaaacttaataacatagaacaatagaaaaggca
ttagatgaatatgaaaggatacttatagaaatagatcggtttatataattatccagaactt
agattaagcgttgatagctctaatagaagaagcacaaaaagttaacgcaaaacttaatacg
actcttggaaaacttgctgggtttattatcttttgggtgattccgagattttggagttacc
gatgagataataagcgaattgaggtctcaaacaaaataccctcattttattaacaactt
20 caagatcgtaagccttatcaattatctgctgatgttgaaaaagtattagctacattaaca
ccaacattgagaagtcggtttgaattgtatggtactacaaagagtttggtatattaatttt
gaatcggttgattatgaggggtgttacctatccattggattatgcaacatttgaaaatgaa
tatgaagatcatccatctcctgaatttagacgtaaaagtttagagcttttagtgatgca
ttacgacaatatcaacatacgacgcccgaacataataatgcaagccaacaagaaaag
25 attgaagcggatttacgaggatatgattctgttattgattatctactacaagatcaagaa
gtaacaaaagatatgttcgatagacaaattgatgtcattatgagtgatttagccccagt
atgcaaaagtatgcaaaaattattcaacgtgtacataaacctggataaaaatgcgatttgag
gatttaaaaatttcaatagaccctaactttgaaccagaaatatcaattgaagaatcgaaa
aaatacattttatggagcgctcaaagtacttgggtgatgattatgtcaaaatgtagagct
30 gcctatgattaccggttggtattttgtctcagaataaaggaaaagatactggagcatat
tgtgcaagtcacatacattacacattcatatgtattttatcatggactgggaaaatggct
gaaacattcgttcttgccgatgaattaggacatgcaggtcattttacattagcgcagaat
catcaaaaattgttggaatctgaagcgctctatgtattttgtagaagcaccttcacaatg
aatgaaatgttgatggcaaaattacttatttaatagtagtaataatcctcgatttaaacgt
35 tgggttatttggttcgattttatctcgaacttattatcataatatggttaccacctttta
gaagcagcttatcaacgtgaagtgtatagccgagtcgacaatggagagtcattaactgcc
ccactgctaaatgaaataatgttgaaacacttataaaacatttttcggtgacactgttgaa
atgacagatggggttgaaatgaatggatgagacaaccacattattatatgggattgtac
tcatatagctactctgctggattgacaattggtacagttgtatcacaatgtatcaagaaa
40 gaaggtcaacctgctgttgatcgctgggttaaaaacgctacaagctggtggtagtcaatct
ccaattgaattggcgcaaatagctggcggttgatattacgactgacgcccctttaaagag
acaattaactatatttcaaaatttagtagatgaattagaagtattaacatatcaataaaaa
gaaattcataa

45 Sequence 1684

MSEGLPLREEVPVKETWDLKDLFTSDQAFYQTLQVQMSLDFNHTYYQKLNNIETIEKA
LDEYERILIEIDRLYNYPELRSLVDTSNEEAQKVNKLNTTSGKLAGLLSFVDSEILELP
DEIISELSQTKYPHFQIKQLQDRKPYQLSADVEKVLATLTPTLRSPFELYGTTKSLDINF
ESFDYEGVTYPLDYATFENEYEDHPSPEFRRKSFRAFSDALRQYQHTTAATYNMVQVQEK
50 IEADLRGYDSVIDYLLQDQEVTKDMFDRQIDVIMSLAPVMQKYAKIIQRVHNLDKMRFE
DLKISIDPNFEPEISIEESKKYIYGALKVLGDDYVKMLESAYDYRWIDFAQNKGKDTGAY
CASPYITHSYVFISWTGKMAETFVLAHELGHAGHFTLAQNHQNLLESEASMYFVEAPSTM
NEMLMANYLFNSSNNPRFKRWVIGSILSRYYHNMVTHLLEAAYQREVYSRVDNGESLTA
PLLNEIMLNTYKAFFGDTVEMTDGVELTWMRQPHYMGLYSYTYSAGLTIGTVVSQCIKK
55 EGQPAVDRWLKTLQAGGSQSPIELAQIAGVDITTDAPLKETINYISNLVDELEVLTYYQIK
ENS*

Sequence 1685

Contig_0640_pos_9870_9118,

is similar to (with p-value 1.0e-31)

>sp:sp|P46339|YQGH_BACSU PROBABLE ABC TRANSPORTER PERMEASE P
ROTEIN IN SODA-COMGA INTERGENIC REGION (ORF72). >gp:gp|D8443
2|BACJH642_159 Bacillus subtilis DNA, 283 Kb region containi
ng skin element. NID: g2627063. >gp:gp|D58414|BACPST_2 Bacil
lus subtilis DNA for homologues of the E. coli pst gene prod
ucts. NID: g903302. >gp:gp|Z99116|BSUB0013_208 Bacillus subt
ilis complete genome (section 13 of 21): from 2395261 to 261
3730. NID: g2634723.
gtgccattatctaaattctttttatctgttacttggaaaccaactggttcatcaccagaa
tttggaaatttgggcactcatcatcggcacattgaaaattactgttattgctacaatcgta
cgagttccgattggttttaggagcagctatctaccttaataatgaatatgcatctgacgttca
cgtagaattcattaaaccaatattagaaattttggctgggattcctacaattgtatttggc
ttctttgctttgacatttgttacacctatattgagaaatctaattcctaacttgggagag
tttaattcaatcagtcgccgtattgttgtgggtataatgattgtccctatgattacaagt
atgagtgaaagatgcaatgtcatctgtacctgataaaattcgtgaaggtgcatttggatta
ggcgcaactaaatttgaagtcgctacaaaagtgttattaccagctgcgacttctggcgtt
gttgcctcaattgtatttaggtatatcaagagcaataggtgaaacaatgacgtttcttta
gctgctggttagttaccacaacatcatctctaagtttaactagttcaattcaaacgatgaca
ggatataattgttgaattgtacaggtgatgcagcatttgggtctgatatttactacagt
atttacgctgtaggttttacacttttcattttcactttaattatgaatttattatcaca
tgatctctaacgattcagagaggagtattaa

Sequence 1686

VPLSKFFLSGTWNPTGSSPEFGIOWALIIGTLKITVIATIVAVPIGLGAAIYLNEYASDRS
RRIIKPILEILAGIPTIVFGFFALTFVTPILRNLI PNLFGEFNSISPGIVVGIMIVMITS
MSEDAMSSVPDKIREGAFGLGATKFEVATKVVLPAATSGVVASIVLGISRAIGETMIVSL
AAGSSPTSSLSTSSIQTMTGYIVEIATGDAAFSGSDIYYSIYAVGFTLFIFTLIMNLLSQ
WISKRFREY*

Sequence 1687

Contig_0640_pos_9116_8232,
is similar to (with p-value 2.0e-27)
>sp:sp|P46340|YQGI_BACSU PROBABLE ABC TRANSPORTER PERMEASE P
ROTEIN IN SODA-COMGA INTERGENIC REGION (ORF73). >gp:gp|D8443
2|BACJH642_160 Bacillus subtilis DNA, 283 Kb region containi
ng skin element. NID: g2627063. >gp:gp|D58414|BACPST_3 Bacil
lus subtilis DNA for homologues of the E. coli pst gene prod
ucts. NID: g903302. >gp:gp|Z99116|BSUB0013_207 Bacillus subt
ilis complete genome (section 13 of 21): from 2395261 to 261
3730. NID: g2634723.
atgtctacacattcaaatactgctaacaaaacattgattgataaagatgccgtagaaaaa
aaaatttcttctcgtgataggaaaaactcggtaacaaatggttatttttattatgtaca
ttaattgggctcattgttttagtagcactattaattcaaactttcgttaaaggggcgga
catctaactcccgaatttttactaatttttcatcttcaacaccagcagatgctggtatt
aaaggggcttttagtaggttctatttggttaatcttaagtattattccaattagtatcatt
ttaggaatagggtacagcaattttattagaagaatacgcaagagacaatatttttacaaa
atcgtaaagggtgagtatatctaatttagctggtgttccctcaattgttttcggtttacta
ggttatacattattttaggcgcggcaggttttaggtaatagcgtgctagccgctgcgctt
acaatgtcactactaatcttgctgttattatcggtgctagtcaggaagctatcagagca
gttcctagttcagtcagagaagcatcatatggtcttggtgctaataatggcaacaatt
agaagagttgttttacctgcagcattacctggtattttaacaggtttcattttatcttta
tcacgcgcattaggagaacacagcaccacttgaatgatagggtatcctacgatactttta
gcaacaccaagcttgattactcgaccaatttctcgtgttaccactcaaatattatcatg
ggcaaaaatgcctcaagcagaatttcaaaaacgttgcatcagcaggtattatcggttctact
cgttatcttattattgatgaacactgtagcgatacttcttcgtaa

Sequence 1688

MSTHSNTANKTLIDKDAVEKKISSRDRKNSVKNWFLFLCTLIGLIVLVALLIQTFVKGAG

HLTPEFFTNSSTPADAGIKGALVGSIWLLISIIPIISIIILGIGTAIYLEEYARDNIFTQ
 IVKVSISNLAGVPSIVFGLGYTLFVGAAGLGNVLAALTMSELLLPVIVASQEAIRA
 VPSSVREASYGLGANKWQTIIRRVVLPALPGILTGFILSLSRALGETAPLVMIGIPTILL
 ATPSGLLDQFLCVTNSNLYMGKNASSRIKRCISRYRSTRYLIIDEHCSDTSS*

5

Sequence 1689

Contig_0640_pos_8117_7242,

is similar to (with p-value 7.0e-95)

>sp:sp|P46342|YQ GK_BACSU HYPOTHETICAL ABC TRANSPORTER ATP-BI
 10 NDING PROTEIN IN SODA-COMGA INTERGENIC REGION (ORF75). >gp:g
 p|D84432|BACJH642_162 Bacillus subtilis DNA, 283 Kb region c
 ontaining skin element. NID: g2627063. >gp:gp|D58414|BACPST_
 5 Bacillus subtilis DNA for homologues of the E. coli pst ge
 ne products. NID: g903302. >gp:gp|Z99116|BSUB0013_205 Bacill
 15 us subtilis complete genome (section 13 of 21): from 2395261
 to 2613730. NID: g2634723.

atggctaattcacaaagtagcagaaaaagagaaactagacgcacaaacaaataatcaagac
 tcagttgccacaatagtaactactgaaaacaataagaaaaatacaattccagacagtga
 aagaagattgtttattcaactcaaaatctagatttatggtaggagaaaatcatgcgtta
 20 caaacattaatttagatataattggaaaataatgtaactgcaataatcgacattctgga
 tgtgtgtaaatctacatacatcaaagcttttaaatagaatggtagtccatctgtg
 aaaactgcaggtaaaattttgtatcgtgaccaaataatattgatgcaaagtattctaaa
 gaaaagctacgtactaacgttggatggctttcaacaacctaaccattccctaagtca
 atttatgataatattacttatggcctaagactcacggtattaaaaacaaaaaaattcta
 25 gatgaaatcgtagaaaaatcattacgtggcgctgcaatatgggatgaattaaaageta
 ttgcatacaaatgcttatggattatcaggtggacaacaacaacgtgtttgtatagctaga
 tgtttagcaattgaaccagatgtcatttttaattggatgaacctacatcagcatttagatcct
 atttctacgtttaagagttgaagaactgtgacaagaattaaaagaaaattactcaattatc
 atggttacgcacaacatgcaacaagctgcgcgtgtttcagataaaaactgctttcttcta
 30 aatggatatgtcaatgaatatgatgataactgataaaatctttcaaactcctgccgacaaa
 caaactgaagattatatatctggctggttttgataa

Sequence 1690

35 MANSQVAEKEKLDQNTNNQDSVATIVTTENNKKNTIPDSEKKIVYSTQNLDLWYGENHAL
 QNINLDILENNVTAIIGPSGCGKSTYIKALNRMVELVPSVKTACKILYRDQNI F DAKYSK
 EKLRTNVGMVFQQPNPFPKSIYDNITYGPKTHGKKNKILDEIVEKSLRGAAIWDELKDR
 LHTNAYGLSGGQQQRVCIRCLAIEPDVILMDEPTSLDPISTLRVEELVQELKENYSII
 MVTHNMQQAARVSDKTAFFLNGYVNEYDDTDKIFSNPADKQTEDYISGRFG*

40 Sequence 1691

Contig_0640_pos_4420_4100,

putative peptide of unknown function

atgcaaaactatttggtagtaccattcttggtagtattttattaactgtattttacaca
 cttgtattaaataaatggtccagctctgcaatcattactttttagtagtttttagcagtactt
 45 gccttttttataccaaattttcaaaacatttcatatcaaccactgcttgatattgcagga
 ttcttaggcataatgagcttaatacagaactttcttatttggtagttttctagaaactgg
 agaaaaaatcgtagaaaaataaattggaaaaagagattcgcaaatatgatgatgaagag
 tcacttcgtcgtcataaataa

50 Sequence 1692

MPNYLWITILGMILLTVFYTLVLNKFQSAIITFVVLAVLAFFIPNFQNISYQPLLGYAG
 FLGIMSLIISFLIWFYFSRNRKRRRIKLEKEIRKYDDEESLRHK*

Sequence 1693

55 Contig_0640_pos_3939_2920,

is similar to (with p-value 0.0e+00)

>gp:gp|AF076684|AF076684_1 Staphylococcus aureus oligopeptid
 e transporter putative membrane permease domain (opp-2B), ol
 igopeptide transporter putative membrane permease domain (op

p-2C), oligopeptide transporter putative ATPase domain (opp-2D), and oligopeptide transporter putative ATPase domain (opp-2F) genes, complete cds. NID: g3800824.

```

5  gtgaaaggatgccacatatgtttaaaatgataattttataaaactttcacaaatgattgtc
   gtactatttatattaactacaatcacatttatattaatgaaactctctccaggtaatcct
   gtagacaaaatttacatcttgatatttcgcatgtatctaagagcaaatagaaacgaca
   gagaataagcttggtctaaataatcctatttttattcaatggtgggactgggttaaatcaa
   ttgtttcattttgatttaggaacaagttatcaaacaagcgagcctgtaattagggaaata
   gcaattatcttggtcctacacttattattacttttggtagcgttatagtgtcattagtt
10  atttctataaccattagggtattatagcagcggtttactaccataaaaatttgggataggata
   atccgtgtttatgacattatccgtaagcctaccatcattttttatcggtcttatctta
   ttatatatatattagcttgaagttgaatattttaccaacttcagatgaggggcgtttcggtt
   tcatatattttaccataaattaccatgagtattggaatgtgtgcttattatattcgattt
   attcggttctactttattagaacaatatcaaacacctatagttgaatcgctctcgctcaga
15  ggtatgcccgaagatatatactttttcaagatatccttaaacctacgatactaccaatc
   atacctctattaggattatccattggtagtttgataggtggaacagtagtcattgaaaat
   ttatttgatattcctgggttaggctattttttagttgacagtataaagtcgagagattat
   ccagtcattcaagggtgtgtgtattttattggtttctttgtagtattataaacacaatt
   cgagatttactttcattacttatcgatcctaaacaacggttatgctattactcagaaagaa
20  acatcaaagtttaaattggtttaattcacatagaaaagaaggtcgtaacgatgaagttaa

```

Sequence 1694

```

25  VKGCQHMFKMIYKLSQMIVVLFILTTITFILMKLSPGNPVDKILHLDISHVSNEQIETT
   ENKLGLNNPFIQWWDWLNQLFHFDLGTSYQTSEPVIREIANYLGPTLIITFGTLIVSLV
   ISIPLGIIAAVYYHKIWDRIIRVMTSLSVSLPSFFIGLILLYIFSLKLNILPTSDEGREV
   SYILPIITMSIGMCAYYIRFIRSTLLEQYQTPIVESSRLRGMPERYILFQDILKPTILPI
   IPLLGLSIGSLIGGTVVIIENLFDIPGLGYFLVDSIKSRDYPVIQGCVLFIGFFVVIINTI
   ADLLSLIIDPKQRYAITQKETSFKFWFNSHRKEGRNDEV*
30

```

Sequence 1695

Contig_0640_pos_2795_2076,

is similar to (with p-value 2.0e-83)

```

35  >gp:gp|AF076684|AF076684_2 Staphylococcus aureus oligopeptid
   e transporter putative membrane permease domain (opp-2B), ol
   i gopeptide transporter putative membrane permease domain (op
   p-2C), oligopeptide transporter putative ATPase domain (opp-
   2D), and oligopeptide transporter putative ATPase domain (op
   p-2F) genes, complete cds. NID: g3800824.
40  atgagtgatgaatcatttacttggcacagatgattacggacgtgatttattcagtcgttta
   gtcgtgggctctcggtgcaacattgtttgttacctacttactttacttttactgtagtg
   gttggagtaccttttagggttacttgcaggctataaaaaagggttgattgatacgattatc
   atgcgaattattgatataggattaagcataaccagaattcggtattatgattgccttagca
   agtttttttcatcctagtcctttggaatttagtaataagctattacaatcataaaatggatg
45  aattatactcgcggtgacaagagggttggtcaataccgaaatgaatcaatcgatatatacag
   atggcacaattttttaatgtctcaactttgaatatcttattttaaacacttattaccaaaa
   gttttaccatctatatttgttattatgatagttgattttggaaaaatcattttatacatt
   agttcattatcatttttaggttttaggtgcacaaccaccatctccagagtggggggcaatg
   ttacaagcagggcggtgaatttattacttcacatcctatcatgattatcgctccagcatct
50  ttgatatacaggtacaatattgatatttaatttaactgggtgatgctgtaagagatcgttta
   ttagaacaagagggtgtaaaagtgtgaaacttttaacaataaaaaatctaaacatcaatga

```

Sequence 1696

```

55  MSMNHLLGTDYGRDLFSRLVVGSRATLFVTLTLLFTVVVGVLGGLLAGYKKGWIDTII
   MRIIDIGLSIPEFVIMIALASFHPSLWNLVIAITIIKWMNYTRVTRGIVNTEMNQSYIQ
   MAQFFNVSTINILFKHLLPKVLPSIFVIMIVDFGKIILYISSLSFLGLGAQPPSPEWGAM
   LQAGREFITSHPIIMIIAPASLISGTILIFNLTGDAVRDLLEQRGVKVETFNKKSKHQ*

```

Sequence 1697

Contig_0640_pos_1936_1337,

is similar to (with p-value 1.0e-64)

5 >gp:gp|AF076684|AF076684_3 Staphylococcus aureus oligopeptid
e transporter putative membrane permease domain (opp-2B), ol
igopeptide transporter putative membrane permease domain (op
p-2C), oligopeptide transporter putative ATPase domain (opp-
2D), and oligopeptide transporter putative ATPase domain (op
10 p-2F) genes, complete cds. NID: g3800824.
atgagtttcgatgaatttaaaatgcaagggtcaaaatacttctggtatcaagcaactttta
ggtaaacatatcggtatatctctcaaaattatgctcaaagttttaatgaatatactcgt
ttggataaacaacttatagctatatatcggtatcatttttaatgtttctaaggataatgca
ttgaaaaagataaaaaaagctttaacttgggttaacttaaatgatgaatcaatcattaat
15 aaatatagtttccaactttcaggaggacaattagagcgagtttaatttgctagcggtttta
atggttagatccagaattaattattgcagatgaacctgttgcatcttttagatgtagtgaac
ggatcatcaataatgcaactccttcaacacattgttaaagatcatcataatactgtatta
cttatcactcataacatgaatcatgtcctcaaataatgctgatttttaattgaatgaga
aatggcatgatgattgaatctggagaaatagacaattatttaaccatcatcttcat
20 cggtatacagaacaattattaaactatagaagcaagctgcaaaaggaggacaacatctaa

Sequence 1698

MSFDEFKMQGQNTSGIKQLLGKHIGYISQNYAQSFNEYTRLKQLIAIYRYHFNVSKDNA
25 LKKIKKALTWVNLNDESIINKYSFQLSGGQLERVNIASVLMMLDPELIADPEVASLOVVN
GHQIMQLLQHIVKDHNTVLLITHNMNHVLKYADYFNVMRNGMMIESGEIDKLFNHHHLII
RYTEQLLNYSKQLQKEDNI*

Sequence 1699

Contig_0640_pos_1334_636,

is similar to (with p-value 9.0e-72)

>gp:gp|AF076684|AF076684_4 Staphylococcus aureus oligopeptid
e transporter putative membrane permease domain (opp-2B), ol
igopeptide transporter putative membrane permease domain (op
35 p-2C), oligopeptide transporter putative ATPase domain (opp-
2D), and oligopeptide transporter putative ATPase domain (op
p-2F) genes, complete cds. NID: g3800824.
atgattcaatttgatcatgtagattattcatatcatcgaaaacagcctgttttaaaagat
attaatataagattattcaacgtggtgaaaaaataggggttttaggggaaagcgggtgctgga
40 aaaagtactattgggttcttaatttaggtcaattaaagccaacaaaaggaaaaataagt
atcgattcaggaaaagggttctacctatttttcaacatgcgacagaaaagtttgatcgtaa
ttcacgattgaacagtcctttgagagagccacttttattttatcgacaattaatacgacaa
aatatcaaaaatatcattcttaactatttaattgaatttaatttgcctacagatctaata
acaaagtttcctcaagaggttaagtgggtgggcaactacaaagattaaatattatagttct
45 ctcttagcacaaccagatatattggtttgtgatgaaataaacttcgaacttagacgtcatg
gccgaacaaaatgtaataatattttacttaacgaaaaaacattcaaaataaaacacta
atcgatcatctcgcatgatttatctgttttacaaaagggttaacgaataggataatagttatc
aaagacggtcaaatagtagatgatttttaaaagtaaaagatttttagccataaaagacat
ccatatacaaaactattaattcaaacgtatgaatattga

50

Sequence 1700

MIQFDHVDYSYHRKQPVLKDINISIQRGEKIGVLGESGAGKSTIGSLILGQLKPTKGKIS
IDSGKVLPIFQHATESFDRQFTIEQSLREPLLFYRQLIRQNIKNIIILNYLIEFNLSTDLI
TKFPQEVSGQLQRLNIIRSLLAQPDILVCDEITSNLDVMAEQNVINILLNEKNIQNKTL
55 IVISHDLVSLQRLTNRIIVIKDGQIVDDFKSKDLFSKRRHPYTKLLIQTYEY*

Sequence 1701

Contig_0641_pos_551_1597,

is similar to (with p-value 1.0e-34)

>sp:sp|P09122|DP3X_BACSU DNA POLYMERASE III SUBUNITS GAMMA A
ND TAU (EC 2.7.7.7).
atggatcaagcaatagcgtttggagacgaacgacttactttacaagatgctttaaatggt
acaggtagtggtgatgaagcggcattaaatgagttatttaattgacattgtaaaaagtgat
5 gttaaagccgcatttaatagatatcatcattttatttcagaaggtaaaagaagtcaacaga
ctcattaatgatatgatttactttgttagagatacaattatgaataaaacgtctaacgaa
tccggttcattttgaatcacttattcatttcgacttagatatgttatacaggatgatagat
atcatcaatgatatactagtagtccattaggttcagtgtaaatcaaagtgttcattttgaa
gtgttgctagttaaacttgcagaaatgattaagacacagcctcaaactgtacaaaatgta
10 gcaacagcatcggtagctaatagaaccagataatgagatgttattacaacgttttagaaciaa
cttgaaaaatgagcttaaaaccttaaaagaacaagggatcaaaactaataaagttagtcaa
caacctaaagaaccaacacgtacgattcaacgatctaaaaatacgttttctatgcaaciaa
atagcgaaagtattagacaaagcaaacaaagatgatatacaattgttgagaaccattgg
caagaagtgattgatcatgcaaaaagtaataataaaaagtccttagtaagtttgctactg
15 aattcagaaccagtagcagctagtagaatcatgtgttagttaaatttgatgaagaaatt
cattgtgaaatagtaataaagatgatgaaaagagaaacaatattgaaagtgtagtttgt
aatatagttataaaaactgtcaaagtagttggagtgcgggctgaccaatggctgagagtg
agagcagagtacttacaaaatcgtaacaccaatgaacacatcaaagcgaaaaacaaagc
acacaacagtcctcaacaatagatatgtctcaaaaagctaaagacttatttggtgaggaa
20 actgtacacttagttgatgaagactga

Sequence 1702
MDQAIAFGDERLTLQDALNVTGSDVDEAALNELFNDIVKSDVKAFFNRYHHFISEGKEVNR
LINDMIYFVRDITMKNKTSNESVHFESLIHFDLDMLYRMIDIINDTLVSIRFSVNQSVHFE
25 VLLVKLAEMIKTPQPTVQNVATASVANEPDNEMLLQRLEQLENELKTLKEQGIKTNKVSQ
QPKKPTRTIQRSKNTFSMQQIAKVLDKANKDDIKLLKNHWQEVIDHAKSNDKSLVSLLL
NSEPVAASEDHVLVKFDEEIHCEIVNKDDEKRNNIESVVCNIVNKTVKVVGVPADQWLRV
RAEYLQNRNTNETHQSEKQSTQSQQIDIAQKAKDLFGEETVHLVDED*

30 Sequence 1703
Contig_0641_pos_2010_2606,
is similar to (with p-value 3.0e-87)
>sp:sp|P24277|RECR_BACSU RECOMBINATION PROTEIN RECR. >gp:gp|
D26185|BAC180K_85 B. subtilis DNA, 180 kilobase region of re
35 plication origin. NID: g467326. >gp:gp|X17014|BSRECM_3 Bacil
lus subtilis dnaZX and recR genes and two unidentified readi
ng frames. NID: g453238. >gp:gp|Z99104|BSUB0001_21 Bacillus
subtilis complete genome (section 1 of 21): from 1 to 213080
. NID: g2632267.
40 atgcattatccagaacctatatcaaagcttatcgatagttttatgaaactgccaggcatt
ggaccaaaagacggctcaacgtctggtttttcacttttagatatgaaagaagacgatgtt
gttaagtttgctaaagcactagttgatgttaaaagagaacttacctattgtagtgtttgt
gggcatattacagaaaatgatccttggttatatatgtgaagataaacagcgagatcgttct
gtcatatgtgttagttgaagatgacaaggatgtcatagcaatggaaaaaatgcgtgaatat
45 aaaggtttatatcacgtgcttcattggttcgatttcaccaatggatggtattggcgctgaa
gacatcaatataacctgcattagttgaacgcctcaaaaacgatgaggtgaaagagcttata
ttagctatgaatcctaacctagaaggcgagtctactgcaatgtatatatctaggttggtt
aaaccaattgggattaaagtcacaagactggcacaaggtttatctgtaggcgcgattta
gaatatgctgatgaagtgactttatctaaagcaattgcaggtagaacggaaatgtaa
50

Sequence 1704
MHYPEPISKLIDFSMKLPGIGPKTAQRLAFHTLDMKEDDVVKFAKALVDVKRELYCSVC
GHITENDPCYICEDKQDRSVICVVEDDKDVIAMEKMREYKGLYHVLHGSISPMDGIGPE
DINIPALVERLKNDEVKELILAMNPNEGESTAMYISRLVKPIGKIVTRLAQGLSVGGDL
55 EYADEVTLSKAIAGRTEM*

Sequence 1705
Contig_0641_pos_4554_4898,
putative peptide of unknown function

gtgacaaaccggaggaaggtggggatgacgtcaaatacatcatgcccttatgatttgggc
tacacacgtgctacaatggacaatacaaaaggcgagcgaactgcgaggtcaagcaaatcc
cataaagtgttctcagttcggattgtagtctgcaactcgactatatgaagctggaatcg
ctagtaatcgtagatcagcatgctacggtgaatacgttcccgggtcttgtagacacccgcc
5 cgtcacaccacgagagtttgtaacaccgaagccggtggagtaaccatttgagctagcc
gtcgaagggtgggacaaatgattgggggtgaagtcgtaacaaggtag

Sequence 1706

VTNRRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSDDCSLQLDYMKLES
10 LVIVDQHAHVNTFPGLVHTARHTTRVCNTRSRWSNHLELAVEGGTNDWGEVVTR*

Sequence 1707

Contig_0641_pos_3048_2737,
putative peptide of unknown function
15 gtgttatacagaaaaatggttgctaaaaataaaacttaaacgcctatacaaacctattca
caaactttatTTTTGTTGTTCAATTTAAAATATTTTACTATTTACTATAATCCTTGTT
TTTTATCCAAATACCTTACATCATCCTTGGGTAAGAAGGATTGATTCTTATCCTCATT
aataataaatgtaattataaaagcctttccgtgaactcaataagtcctgaattctaaaaag
cgaaacagaaatcttatcatattcttctttgtttcaatccattaccgaacccaacttgc
20 tttgtctgttga

Sequence 1708

VLYRKMVAKIKLRLYKPYSQTLFLLFILKYFYYFTIILVFYISKYFTSSLGKEGLILILI
25 NNKCNYSLSVNSISLNSKKRNRNLIIFFFVSIHYRTPTCFVC*

Sequence 1709

Contig_0642_pos_3898_4338,
putative peptide of unknown function
30 atgggtgaagttaaacatttacaatacaattacaaaacagatgagcttttcgctgatttc
agagaatttggaacaagaacctatatatgattgaggaggttaaagggacaaatgattgat
gcaagttccgattctcctttttatggaattttgttggttaataaattagttgcgagaatg
gcattatttgataaaggagaagtagaagaacatttccctaataagcaatgactatatt
cttctttggaaattagaagtagtagatagctaccaaagcgtggatagctaaagcaacta
ttaaactttgctaaagaaaaataaaaaaccaattaaagctattgctagaataaactcaaaa
35 gaattctttttaaagcagggttttaaagatgtagaaacaaaaatcctgagggggcacgac
atcttaatatggaatccataa

Sequence 1710

MGEVKHLQINYKTDELFDREFGNKNLYMIEELKGQIMIDASSDSPFYGIFVGNKLVARM
40 ALLDKGEVEETYPNSNDYILLWKLEVLDTYQRRGYAKQLLNFAKENKKPIKAIARNNSK
EFFLKQGFKDVKETKNPEGHDIWNP*

Sequence 1711

Contig_0642_pos_3719_2106,
45 is similar to (with p-value 9.0e-71)
>sp:sp|P55342|YLLA_BACSU HYPOTHETICAL 62.6 KD PROTEIN IN FTS
L 5'REGION. >gp:gp|Z68230|BSYLLSPO_1 B.subtilis yllA, yllB,
yllC, ftsL, pbpB and spoVD genes. NID: gll22757.
atgaagtgaatacgttaaaagttaactgaacaggatcagttcattaataaaaaataaaaaat
50 agtgaatctcaaattacatctttttatgaatatgatgcagcgaagaaaggaaagcttttat
agaagattaaaaacacctaataatggaagggaatttcatttatcacgagtgattaaatct
tacatgaatgaattaaagttaacacatcagcagctgaataacatagatgctttagctgat
gggtgcaaaagttagtgattgggtggacaacaagctgggtctttttgggggtccactgtatacg
tttcataagatttttctcaattataactttaagtcgtcaactttctgaggaatatgatact
cctattgtacctgtgttttggttgaggtgagggaccatgattttgaagaggtgaatcat
55 acgtatgcattcaataataaagaaactaccttgaaaaaagtaaagtatcatacagtgaca
ccaccagatagtaattgtttcaagatatctcctgataagaacgaattaaaagcatcatta
aatcacttttttaaagaaatgaaggaaactgtacatactcaagatgtttatcaaatgtgc
gtcaatattattaatcaatttgattcatggattgatatttttaagggttgatacatgaa

gtgtttaaggattatgggattttacttattgatgctcaataccctgaattaagacagatg
 gagaaaccgttggttaagagatattagaaaaggaggaatcaagtcgatcaatcttttcgt
 gaaactcagatacgaacaactcaacaacaacttccatcaatgatacaaacagagacaaac
 acacattttatccatgaagacggaatgagacagcttttaattttgatggcacttat
 5 tttaaactgaataaaaactgagaacggttacacgaaacaaaatttattagatattatagaa
 agagagcctgaaagaatttctaataatgttggttactcgtccagttgtagaggaatggttg
 tttacacagtagcattttatcggaggtccaagcgaaatcaaataattgggcagaattaaaa
 ggtgtttttgatagcgttaaatgtagaaatgcctattgttatgccagattaagaatcacg
 tatttgtatgctagaactaaaaagttattaaaacaataataatttatcgatagagtctgtc
 10 attgctaattggagtagaacaggaacgtcaacggtttgttcgtgaaaaagcatcaaataat
 tttataaatgaagtagaagaaatgaaaattcagcaacaagaactttataacaatttattc
 acctatgtggaaataatcatgacaaccaacttcttttagaaaaaaataatcaaattcat
 ctcaatcagtagcattattttaatcaaacgggtacttactgaatattgaaagagaaaatgat
 ataagtatgcgacagtttagagaaattagtgaacactccatccaatgggtggtctacaa
 15 gaaagagtttggaatccacttcaaattatgaatgattttgggatagatgtgttcagtcctc
 ccacctatccaccactttcttactcgtttgatcatttgattataaatccttga

Sequence 1712

MKCNTLKLTEQDQFINKIKNSEQITSFYEYDAAKKESFYRLKTPNNGREFHLSRVIKS
 20 YMNELKLTHQQLNNIDALADGAKVVIGGQAGLFGGPLYTFHKIFSIIITLSRQLSEEDYT
 PIVPVFWIAGEDHDFEEVNHTYAFNNKETTLKKVKYHTMTPDSNVSRYTPDKNELKASL
 NHFFKEMKETVHTQDVYQMCVNIINQFDSWIDIFKGLIHEVFKDYGILLIDAQYPELRQM
 EKPLFKEILEKRNVQDQSFRETQIRTTQQQLPSMIQTETNTHLFIHEDGMRQLLNFDGTY
 FKLNKTEKRYTKQNLLDIIEREPERISNNVTRPVVEEWFNTVAFIGGPSEIKYWAEK
 25 GFVDTLNVEMPIVMPRLRITYLYARTKKLLKQYNLSIESVIANGVEQERQRFVREKASNN
 GINEVEEMKIQOQELYNLFTYVENNHNDQLLLEKNNQIHLNQYDYLIKRYLLNIEREND
 ISMRQFREISETLHPMGLQERVWNPLQIMNDFGIDVFSPTYPPLSYSDHLIINP*

Sequence 1713

Contig_0642_pos_1962_1531,
 is similar to (with p-value 8.0e-70)
 >sp:sp|O07319|YLLB_STAAU HYPOTHETICAL 17.4 KD PROTEIN. >gp:g
 p|U94706|SAU94706_1 Staphylococcus aureus strain ATCC 8325-4
 cell wall/cell division gene cluster, yllB, yllC, yllD, pbp
 35 A, mraY, murD, divlB, ftsA and ftsZ genes, complete cds. NID
 : g2149889.
 atgttcattgggagaattcgatcatcaattggatacaaaaaggacgtatgattataccgtcc
 aaatttcgttatgatctaaatgaacggtttattatcacaaagaggccttgataaatgttta
 tttggttacactctagaagagtggcagcaaatgaagagaagatgaaaaccttacctatg
 40 acaaaaaagacgcgcgtaaatttatgcgtatgttctctcaggtgctgtagaagtagaa
 ttagataaaacaagggcgtattaatattccgcaaaatttaagaaaatatgccaatttaagt
 aaggaatgtacagtaattggtgtctcaaatcgtatagagatttgggacagagaaaacttgg
 aatgatttctatgatgaatctgaagaaagtttcgaagacattgctgaagatttaatagat
 tttgatttttaa

Sequence 1714

MFMGEFDHQLDTKGRMIIPSKFRYDLNERFIITRGLDKCLFGYTLIEWQQIEEKMKTLP
 TKKDARKFMRMFFSGAVEVELDKQGRINIPQNLRYANLSKECTVIGVSNRIEWDRETW
 NDFYDESEESFEDIAEDLIDFDF*

Sequence 1715

Contig_0642_pos_1495_581,
 is similar to (with p-value 0.0e+00)
 >gp:gp|U94706|SAU94706_2 Staphylococcus aureus strain ATCC 8
 325-4 cell wall/cell division gene cluster, yllB, yllC, yllD
 55 , pbpA, mraY, murD, divlB, ftsA and ftsZ genes, complete cds
 . NID: g2149889.
 atgttaaacgaaaccattgattattttaaatattaaagaagatggtgtgtatgttgactgt
 acgttgggtggagcaggacatgccctctatttacttaataatgataaagtaga

cttattgcgattgatcaagatttaacagccatagaaaatgcgaaagaagttttaaagaa
catttgcacaaagtcacttttgttcataacaactttcgagaattaacaaatattttaaat
gaattagaaattgaaaaagtagatggtatttattatgacttaggtgttcaagccgcaa
ttggatgtgcctgaaagaggcttttagttatcacaaatgatgcgaaactagatatgcgaatg
5 gatcaaacaacacactttctgcgtatgaagtagttaatcaatggtcttatgaagcatta
gttaggattttcttctcggttacggtgaagagaaattttctaaacaaattgcacgcagaatt
gaagcccatcgagaacaacaacctatagaaacaactttagaactagtgtgatgtcattaaa
gaaggcataccagcgaaagcaagacgaaaaggggacatcctgcgaaacgcgtgttccaa
gctattcgaaattgctgtgaatgatgagttatcagcttttgaagattcagttgagcaagcc
10 attgaatgtgtgaaggtcgagggtagaatttcagttattactttccactctttggaagat
cgtttgtgtaacaaattttccaagagtttgagaaaggtccagacgtaccaagaggtctc
ccggttattcctgaaagcatatacacctaagttaaaacgagtaaatcgtaaacgattacc
gctactgatgacgattttaaacgaaaacaatcgagcacgtagcgccaagttacgcgtagca
gaaatattaaaataa

15

Sequence 1716

MLNETIDYLNLIKEDGVYVDCITLGGAGHALYLLNQLNDKGRLIAIDQDLTAIENAKEVLKE
HLHKVTFVHNNFRELTNINLELEIEKVDGIYYDLGVSSPQLDVPERGF SYHNDAKLDMRM
DQTQSLSAYEVVNQWSYEALVRIFFRYGEKFSKQIARRIEAHREQPIETTLELVDVIK
20 EGIPAKARRKGGHPAKRVFQAIRIAVNDELSAFEDSVEQAIECVKVGGRISVITFHSLED
RLCKQIFQEFEKGPDPVPRGLPVIPEAYTPKLKRVNRKPITATDDDLNENNRARS AKLRVA
EILK*

25 Sequence 1717

Contig_0645_pos_1187_1612,

putative peptide of unknown function

atgactaatcaaaaccaaattgaacaacgattctataatattattgaaaatgctgatcaa
tattatctctacttatatgattggcgtgtgattgaaagtgaatctaagagagcgacttt
aaagacgttgatatatggattaactttgaagatgaagcggttaatagatgaatataattgt
30 gtgattgctaaggttgatgatgaaggtggcaatattaatcatatgatttctcaaaactta
cgtcacaaatatgtttggtctacaccgttcttgatgagagttgaggagaacttttagacca
tacctgcacatttatggaacatgactttaaaaaggccattgaaaatcagaatgaaatat
aatgcagtatatatttagttgaaatatataaaaaagataaaataaaataaaaggcgtagcaca
acttaa

35

Sequence 1718

MTNQNIQERFYNIENADQYYLYLDWRVIESES NESDFKDVDIWINFEDEALIDEYIC
VIAKVDDEGNNINHMISQNLRHKYVWSTPFLMRVEENFRPYMHIMEHDFKKGPLKIRMKY
NAVYLVEIYKKDKINKRRSTT*

40

Sequence 1719

Contig_0645_pos_3058_3531,

putative peptide of unknown function

atgactgctaattgattggatagaccggttgaattaatttcgcatcctgaaggcggatat
45 tttaaagaacaatgagagggcgagcggtaaaaggaagagcgctcttttagcagtatttatttt
ctattaacacagcgcgatatatcacattttcatagaatagatgcagatgaagtatggtat
tatcatgctggtcgagcgttgaagattcatatgataaccctaaaggatgaatatacact
gttaaattaggtagagatatagattgtggtgagtggttacaatatgtgtgccaaaaggg
acaatttttgcttctactttagatagtgagagggatatagcttagttggatgtatgtgt
50 caacctggatttgaatacagcattttgaacttttaacacaagaatatttgattcgtcaa
tatccacaatatgaaagcataataaaaaagattagctatatctcaagaagatttaa

Sequence 1720

MTANDWIDRLELISHPEGGYFKETMRGDGKGRASFSSIYFLLTQRDISHFHRIDAEVWY
55 YHAGQTLKIHMITPKGEYHTVKLGRDIDCGECLQYCVPKGTIFASTLDSAEGYSLVGCMC
QPGFEYEHFELLTQEYLIRQYPQYESI IKRLAISQED*

Sequence 1721

Contig_0645_pos_6262_6861,

putative peptide of unknown function
 gtgactgtgaaagtaaaatataatagacaaacgtcactggcgtcgccctggtagagagagaa
 tatacagaggttaaagtttaataataatagatttaaaggtattataggcttggtcacgatg
 aaaaaggttcgtgagccttttagaggtgacggtagttggacaaaacataatagttgcagat
 5 gacaattataaaatgggttacagatactaccggataaagaagcggatagtatgactgtcatg
 tttgataataaagggaatccattagaatactactttgacattaataaaagaacattact
 caaaaaggaaatgcacgtaccatcgattttatgtttagacgttctcgttctacctaattggt
 gagtatgaacttgtagatgaagatgatttaattgtacgcactacaaaataaacaatttca
 10 aagaagcaatatcatgaggcatatattatcgcccatcaattaatgattgaaatagaagat
 aatttttcagaaatacaagataaggttatgcgttggttatcataaaatcaatcataaagca
 cagaaaatgaaacataaacgtccctataaagctaaaaagaatcacaccgacgacattaa

Sequence 1722
 15 VTVKVKYIDKRHWRRRLVEREYTEVKVNNNRFKGIIGLVTMKKVREPLEVTVVGQNIIVAD
 DNYKWLQILPDKKRYSMVTFDNKGNPLEYYFDINIKNITQKGNARTIDLCLDVLVLPNG
 EYELVDEDDLMYALQNKQISKQYHEAYIIAHLMIIEIDNFSEIQDKVMRCYHKINHKA
 QMKHKRPYKAKKKSHRRH*

20 Sequence 1723
 Contig_0645_pos_711_385,
 putative peptide of unknown function
 gtgaaagatgatttaaaccgatgattttgaagattcttttagagtatttgagccattagat
 catgatgcataatattgtgaggttaaactttactggtgaaaatacgactgagcctatcata
 25 tcttatatgacgacgacgcataacatagatgtgaatattcttgaagcagatattaagaat
 actaaaaacgggttcggttgatttttagttattcacataccacatataagtgaagaacat
 ttcaagcaatttaaaccataatcttcacacaaaagctaattcttttagtaggtatggctgg
 ggaaagagatttaacgaaaacacctga

30 Sequence 1724
 VKDDLNDFFEDSLEYLEPLDHDAYIVRLNFTGENTTEPIISYMTTTHNIDVNILEADIKN
 TKNSFGFLVIHIPIHISEEHFKQFKHNLHTKANLFSRYGKGKRFNENT*

Sequence 1725
 35 Contig_0645_pos_383_69,
 is similar to (with p-value 5.0e-54)
 >gp:gp|Y14370|SAY14370_2 Staphylococcus aureus RF3, murE, yp
 fP genes. NID: g3256221.
 atgggggagagtagcatgtcgagcagattatgttatttttactccagataatcctgctaac
 40 gatgatcctaaaatgttgacagctgaattagctaaagggtgcaacgcataacaattatata
 gagtttgatgaccgtgcagaaggtattagacacgcgattgatattgctgaaccaggtgat
 acagttgttttggcctcaaaagggtcgagagccttatcaaattatgcctggatgtttaa
 gtccacatcgcgatgacttaattggcttaaaagcagcatatcaaaaatttggtggtgga
 45 cctcttgaggattaa

Sequence 1726
 MGRVACRADYVIFTPDNPNANDPKMLTAEALAKGATHNNYIEFDDRAEGIRHAIDIAEPGD
 TVVLASKGREPYQIMPGHVKVPHRDDLIGLKAAYQKFGGGPLED*

50 Sequence 1727
 Contig_0646_pos_5424_4078,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P16618|HEM1_BACSU GLUTAMYL-TRNA REDUCTASE (EC 1.2.1.-
) (GLUTR). >pir:pir|A35252|A35252 5-aminolevulinate synthase
 55 (EC 2.3.1.37) - Bacillus subtilis >gp:gp|M57676|BACHEMAXC_1
 Bacillus subtilis hemAXCDBL gene cluster. NID: gl43034. >gp
 :gp|Z99118|BSUB0015_82 Bacillus subtilis complete genome (se
 ction 15 of 21): from 2795131 to 3013540. NID: g2635200. >gp
 :gp|Z75208|BS275208_87 B.subtilis genomic sequence 89009bp.

NID: g1769994.

atgcattttgttgcattatagcataaatcatcgaacagctgatgtaacattaagagagcaa
 gttgcttttagagatgatgccttacgattagcacatgaagatttatatgaaactaaagca
 attttagaaaatgtcattttatctacatgtaatcgtactgaagtatacgtattgttgat
 5 caagttcatacaggacgttattatatacaaaagatttttagcgcgctcttttgatttgag
 gtatgatatttaaagatatgtcggaaagttaaagtggggacgatgcagttgaacattta
 ttgctgtcacttctggcttagattcaattgtgcttgggtaaacacaaaatttaggacaa
 atgcgcgatgcatttttcttagcgcaaaatactggtacaactggaacgatttttaacat
 ttatttaaacaagcgattacttttgctaaaaaagcacacagtgaaacagacattgcagat
 10 aatgctgtgagtggttcttatgctgctgttgaaattagctaaaaaggattttggaaaatta
 aaaagtaaacatgctgtcgttattggagcaggggaaatgggtgaattatcactcttaaat
 cttttaggttctggaatttcaaatgtaacaattgttaatcggacattatctaaagctaaa
 attttagcgaaaaaacacaatgttcatatgattcactttcagcattaccatctttatta
 gaaacaacggacatagtaattagttctacaagtgtgaagactatatcatcactaattct
 15 atgggtgaaaaacaatttcagaaactagaaaactggattcattagttctgattgatattgag
 gttccacgagacattgaaccagggttagtgcaattacaaatatttttaattatgatgtt
 gatgatttgaaagatttggttagatgccaaatttaagagaacgtcaattagctgctgaaact
 attgcaggacaaatacctgaggagattgattcacacaacgaatgggttaatatgcttggt
 gttgtacctgtaactcagagctttacgtgaaaaagctatgaatatccaagcagaaactatg
 20 gaaagtattgatcgtaaatgccaagatctctctgaaagagaacgtaaagtcatttcgaaa
 cataaaaaagattatcaatcaaatgttaaaagatcctatcaaacaggctaaggaatta
 agtactgataaaaaagtaattgaaaaattagagctatttcaaacatatttgataattgaa
 gccgaagatcctcgtaaaaaagcaaagttagaaaaagagagtagagcaaaggaaatctta
 gcgcacatgaatatttagttttgaataa

Sequence 1728

MHFVAISINHRTADVTLREQVAFRDDALRLAHEDLYETKAILENVILSTCNRTEVYAI
 QVHTGRYYIQRFILARSFGFEVDDIKDMSEVKVGDDAVEHLLRVTSGLDSIVLGETQILGQ
 MRDAFFLAQNTGTTGTIFNHLFKQAITFAKKAHSETDIADNAVSVSYAAVELAKKVFGKL
 30 KSKHAVVIGAGEMGELSLLNLLSGSISNVTIVNRTLSKAKILAEKHNVSYDSLPSLL
 ETDDIVISSTSAEDYIITNSMVKTISETRKLDLSVLIDIAVPRDIEPGIDAITNIFNYDV
 DDLKDLVDANLRERQLAAETIAGQIPEEIDSHNEWVNMLGVVPVIRALREKAMNIQAETM
 ESIDRKLPDLSEERKRVISKHTKSIINQMLKDPIKQAKELSTDKKSNEKLELFQNI
 AEDPREKAKLEKESRAKEILAHRIFSFE*

Sequence 1729

Contig_0646_pos_3867_3244,
 is similar to (with p-value 1.0e-26)
 >sp:sp|P16645|HEMX_BACSU HEMX PROTEIN. >pir:pir|B35252|B3525
 40 2 hypothetical protein (hemA 3' region) - Bacillus subtilis
 >gp:gp|M57676|BACHEMAXC_2 Bacillus subtilis hemAXCDBL gene c
 luster. NID: g143034. >gp:gp|Z99118|BSUB0015_81 Bacillus sub
 tilis complete genome (section 15 of 21): from 2795131 to 30
 13540. NID: g2635200. >gp:gp|Z75208|BSZ75208_88 B.subtilis g
 45 enomic sequence 89009bp. NID: g1769994.
 gtgccactaggttctatatttgacgttttttctcttaacttgattatttatcaatt
 tcgttgatattgaatttgattaaagtaattgaatttttctgttttttcttaatttgatt
 ggatttatcctaattgagtttaataacttttcagcctgaacattatcaaacgcaaattcaa
 caaattgcagtaattaatgaattattacttgttcataatagcacttgacgtattaagttat
 50 gcattttttgcaatcgcatgtgaaattcattactctacattattcaatatcgaaattta
 aaggagaaaaatttcgatcaaaattactttagaattggtagtggtgctactttagaacc
 atcgttttctattcaacacttgttgcatggattatccttatattaagtacgatttttaggt
 gcacaatggggatcttttcagctcggtaaacaaaattttatagatccaaaagtaatttt
 tcaacaattattaaatttatattggttttatattttcattagaataaaaaaattggata
 55 tcacaaagaaatcttactttaaacttatattttttgtttgtatgattaattta
 ttctttttaactcatttttagataa

Sequence 1730

VPLGSIFDVFFSLTWIIISISLILNLIKVMNFSVFFLNLIIGFILMSLNTFQPEHYQTQIQ

QIAVINELLVHIALAVLSYAFFAIAFVNSLLYIIQYRNLKEKNFDQNYFRIGSVATLET
IVFYSTLVAWIIILSTILGAQWGI FAVGKQIFIDPKVIFSTIINLLYGVYIFIRIKKWI
SQRNLIYFNIILFCLCMINLFFLTHFR*

5 Sequence 1731

Contig_0646_pos_3190_2264,

is similar to (with p-value 0.0e+00)

>gp:gp|U89396|SAU89396_1 Staphylococcus aureus hemCDBL gene
cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
10 III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
: g2589180.

atgcgtaaattaattgttgggtcgcggaagaagtaaattagcgctaacacaaagtcaacaa
tttatagataaaattaaaatttatcgatccgtccttggatattgaaataaaagaaattgta
15 actaaaggcgacaaaattgtagataaaacattatccaaagttggaggtaaaggacttttt
gttaaggaaaatccaaaatgaattatttaataaagagatagatatggcgattcattctcta
aaagatgtaccaagtatgattcctgacggtcttaccttaggatgtattcctgatagagaa
attccttttgatgcctatatagcaaaaaatcatataaccattacaagaattgtctgagggc
agcattgtaggtacaagttctttacgtcgtggcgctcaaattttatcaaaaatcccacat
20 ttaaaaattaaagtggtattcgtggaacattgatactcgattaaaaaaattagagactgaa
tattatgatgctattatattagctgctgctggtattaaaacgcagtggttggtcagataat
attgttacgacttatcttgatcgagatatattactgccagctatagggcagggtgcactt
gggtattgagtgtaggagtgatgacaaagaacttttagatttactatctaaagtacacaat
catgatgtagcacaatgtgtgactgctgaacgtactttctatcagaaatggatggcagt
25 tgtcagggtcctataggtggatagcaacaattgctcaagataaccaaattgaatttaca
ggactgattatgtctccagatggtaaggaaagatatgagcatacagcattgggtactgat
cctgtaaaattgggtatagaagtgagtcagtaacttaaaaaacaaggtgcttatgacata
attaaaaaattaaacgaagcagaataa

30 Sequence 1732

MRKLIVGSRRLKALTQSQQFIDKLKFIDPSLDIEIKEIVTKGDKIVDKQLSKVGCKGLF
VKEIQNELFNKEIDMAIHSLKDVPSMIPDGLTLGCIPDREIPFDAYIAKNHIPLQELSEG
SIVGTSSLLRGAQILSKYPHLKI KWIRGNIDTRLKKLETEDYDAIILAAAGLKRMGWSDN
IVTTYLDRDILLPAIGQALGIECRSDDKELLDLLSKVHNHDVAQCVTARTFLSEMDGS
35 CQVPIGGYATIAQDNQIEFTGLIMSPDGKERYEHTALGTDVPVKGIEVSQVLKKQAYDI
IKKLNEAE*

Sequence 1733

Contig_0646_pos_2230_1550,

40 is similar to (with p-value 5.0e-76)

>gp:gp|U89396|SAU89396_2 Staphylococcus aureus hemCDBL gene
cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
45 : g2589180.

atgaaaccagttatagttatgacgcagacgaatgaagttcatagtcatttagttgatatt
atccataagccttttatccaactaaaacaacttcattttaatgaaaaattgcttgatcat
agctacgactggcttatttttctgctctaaaaacgcagtaaaatacttttatccttattta
aaaaacgtaaaggttaaaaaaggtagctgttataggtgataagacagctcagttattgtaat
50 gaattaggtattagtgctgactttgtgccacgtgatttttctcaagaaggctttttggac
gagtttaagattagcgaacaacatttattgttgccctcaagtgaagaagcacgctcctaaa
ttagttcaacaattgagcaaatataatgaagtcgttaaaattgatttatatagaccagta
ccgaattttaaaaatataagtcgaagtttaagtcctctgttagaaaacatcaaatagacgca
gtgactttttctagttcctctgcagttgaattttatttcaaagaggacaattgtgcctgaa
55 tttgatcattattttgtctatcggttaagcaaaactgctaggaccattttaaattcaataca
tctgtaaaaagtggaataaacaacattagattcacttattgataaaataatagaaagt
agggaacaaaatgaaatttga

Sequence 1734

MKPVIVMTQTNEVHSHLVDIIHKPFIQLKQLHFNEKLLDHSYDWLIFSSKNAVKYFYFYPYL
KNVKVKKVAVIGDKTAQYCNELGISVDFVPRDFSQEGFLDEFKISEQHLLLPSEKARPK
LVQQLSKYNEVVKIDLYRPVPNFKNISQVKSLVRKHQIDAVTFSSSSAVEFYFKEDNVPE
FDHYFAIGKQTARTILKFNTSVKVANKQTLDSLIDKIIESREQNEI*

5

Sequence 1735

Contig_0646_pos_1518_586,

is similar to (with p-value 0.0e+00)

>sp:sp|P50915|HEM2_STAAU DELTA-AMINOLEVULINIC ACID DEHYDRATA
SE (EC 4.2.1.24) (PORPHOBILINOGEN SYNTHASE) (ALAD) (ALADH).
10 >gp:gp|U89396|SAU89396_3 Staphylococcus aureus hemCDBL gene
cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
15 : g2589180.

atgcgtgatttagtaagagaaactcatgttagaaaagaagatttaatatatccaatattt
gtagttgagcaagatgatataaaaaagtgaattaaatcactaccaggcatataccaaatt
agtttaaatttattgcatgaagagattaaaggcatatgatttaggtattagagcaatc
atgttcttcggtgtgccaatgacaaagacgacattggatctggtgcatatgatacaat
20 ggagttgttcaagaagcgacacgaatatctaagaatttatataaggatttacttattgtt
gcagatacttgtctttgtgaatacacagaccacggacactgtggcgttattgacgatcat
acgcatgatgtagacaatgataaatcacttccattacttgtaaaaacagctattttctcaa
gttgaagctggagctgacatcattgctccaagtaatatgatggatggttttgttgctgaa
attcgtgaagccttgatcaagcgggatatacaaatattcctatcatgagttatggtatt
25 aaatagcatcaagcttttttcggtccattcagagatgctgcagattcagcaccttctttt
ggggatagaaaaacctatcaaatggatcctgcaaaccgtttagaggcattaagagaattg
gaaagtgatcttaagaaggttgcgatatgatgatagttaaaccatctttaagttatcta
gatattattagagatgtaaaaataatacgaacgtgccagtcgtagcatacaacgttagt
ggagaatatagtatgacaaaagcagcagcgttaaatggttggatagatgaagagaaaatt
30 gttatggaacaaatgatattctatgaaacgtgcaggtgctgatttaataattacttatttt
gcaaaagatatctgtcgttatttagataaatag

Sequence 1736

MRDLVRETHVRKEDLIYPIFVVEQDDIKSEIKSLPGIYQISLNLHHEEIKEAYDLGIRAI
35 MFFGVPNDKDDIGSGAYDHNGVVQEATRISKNLKDLLIVADTCLCEYTDHGHCGVDDH
THDVDNDKSLPLLKTAISQVEAGADIAPSNMMDGFVAEIREGLDQAGYQNIPIMSYGI
KYASSFFGPFDAADSAPSFGRKTYQMDPANRLEALRELESDLKEGCDMMIVKPSLSYL
DIIRDVKNNTPVFPVAYNVSGEYSMTKAAALNGWIDEKIVMEQMISMKRAGADLIITYF
AKDICRYLDK*

40

Sequence 1737

Contig_0646_pos_0_535,

is similar to (with p-value 6.0e-93)

>gp:gp|U89396|SAU89396_4 Staphylococcus aureus hemCDBL gene
45 cluster: porphobilinogen deaminase (hemC), uroporphyrinogen
III synthase (hemD), d-aminolevulinic acid dehydratase (hemB
) and GSA-1-aminotransferase (hemL) genes, complete cds. NID
: g2589180.

atggagcaagctgagaaattaatgcctggcggtgttaacagtcccgtgaagagcatttaaa
50 tcagtagacacaccagctatttttatggatcatggtgaaggatctaaaatatatgatatt
gatggaaatgaatacattgattatgtgctaagttggggccatttaattctgggacataaaa
aatcaacaagttatatccaaatt'acatgaagcagtagataaaggtacaagcttcggcgct
tcaacacttcaagaaaataaaacttgctgaacttgtagaccgtgtaccttcaattgaa
aaagtaagaatggtttctcagggaactgaagctactttagacacacttcgttttagctagg
55 ggttatacaggacgtaataaaaattataaaatttgaaggggtgttatcatggacacagtgat
tctttattgattaaagcaggatcaggtgttgcaacactaggtttacctgattcaccaggc
gtccctgaaggtatttgctaaaaacactatcacggtgccatataatgatttaTGCC

Sequence 1738

MEQAELKMPGGVNSPVRAFKSVDTPAIFMDHGEKSKIYDIDGNEYIDYVLSWGPLILGHK
 NQQVISKLHEAVDKGTSFGASTLQENKLAELVIDRVPSIEKVRMVSSGTEATLDTLRLAR
 GYTGRNKIIKFEGCYHGHSDSLIKAGSGVATLGLPDSGPVPEGIAKNTITVPYNDLCX

- 5 Sequence 1739
 Contig_0648_pos_234_1580,
 is similar to (with p-value 1.0e-95)
 >sp:sp|Q45493|YKQC_BACSU HYPOTHETICAL 61.5 KD PROTEIN IN ADE
 C-PDHA INTERGENIC REGION. >gp:gp|AF012285|AF012285_29 Bacill
 10 us subtilis mobA-nprE gene region. NID: g3282109. >gp:gp|Z99
 111|BSUB0008_125 Bacillus subtilis complete genome (section
 8 of 21): from 1394791 to 1603020. NID: g2633699.
 atgaaggcccgaaatattaaaaaagaaagtagcttactatactgtaaaccatgattcaatt
 atgagatttaaaaatgttaacgtgagtttctttaatacgcacacatagcattcctgatagc
 15 ttaggcgtatgtattcatacttcgtatggttctatagtttatactggagagtttaagttt
 gatcaaagtttgcattgacattatgctccagacttgaaacgaatggcagaaattggtgat
 gaggggtgtgttcgcattaatcagtgattcaacagaagctgaaaagcctggatataacacg
 cctgaaaatattattgaacatcacatgtatgatgcttttgccaaggttaaaggttagactt
 attgtatcatgctatgcttcaaaacttcgttcgtattcaacaagtgttaacattgcaagt
 20 caacttaattcgtaaaagtgtcatttttaggtcgttcacttgaaagtctggttaacatagca
 cgtaaaatgggatactttgatataccaaaagattttattaatacctattaacgaagtggaa
 aattatcctaaaaatgaagtgtattatttgcacaggtatgcaaggtgaaccagtagaa
 gcattaagtcaaattggctcgcaaaaagcataaaattatgaacatagaagaaggagattca
 atattcctagcaattactgcttcagctaataatggaggttattattgcagatacatttaaat
 25 gagttagtgctgctggagcacatataattccaaacaacaagaaattcatgctgcaagt
 catggttgcattggaagaattgaaaatgatgttaaatattatgaaacctgaatattttgtg
 cctgttcaaggtgaaatttaaaatgcagattgcacatgccaattagcagcagaaaccggt
 gtagcacctgagaaaaattttcttagttgaaaaaggcgacgtgattagttataacggtaaa
 gatattgatttttaaatgaaaaagttcaatcaggtaatatacttattgatgggattggcggt
 30 ggtgacgtaggtaatatcgtattaagagacagacatctatttagccgaagacggtattttt
 attgcggttgtgacatttagatcctaaaaatcgacgtatttgcgcaggacctgaaattcaa
 tcaagaggcttcgtctatgttagagaaagtgaagaacttttgaaagaggctgaagaaaaa
 gtacgtaaaattgtagaggaaggtcttcaagaaaaacgaatagaatggtcagaaatcaag
 caaatatgagagatcaaatcagtaagttactatttgagagtacaaaacgccgtccaatg
 35 attattccagtcataatcgagatctaa

- Sequence 1740
 MKARNIKKKVRYTVDNHSIMRFKNVNVSFNTHSIPDSLGVCIHTSYGSIVYTGEFKF
 DQSLGHGYAPDLKRMAEIGDEGVFALISDSTEAEKPGYNTPENIEHHMYDAFAKVKGR
 40 IVSCYASNFVRIQQVLNIIASQLNRKVSFLGRSLESSFNIARKMGYFDIPKDLLIPINEVE
 NYPKNEVRIIATGMQGEPEVEALSQMARKKHKIMNIEEGDSIFLAITASANMEVIIADTLN
 ELVRAGAHIIIPNNKKIHASSHGCMEELKMMLNIMKPEYFVPVQGEFKMQIAHAKLAAETG
 VAPEKIFLVEKGDVISYNGKDMILNEKVQSGNILIDGIGVDVGNIVLRDRHLLAEDGIF
 IAVVTLDPKNRRRIAAGPEIQSRGFVYVRESEELLKEAEKVRKIVEEGLQEKRIEWSEIK
 45 QNMRDQISKLLFESTKRRPMIIPVISEI*

- Sequence 1741
 Contig_0648_pos_2002_4239,
 is similar to (with p-value 0.0e+00)
 50 >sp:sp|P21458|SP3E_BACSU STAGE III SPORULATION PROTEIN E. >p
 ir:pir|S09411|S09411 spoIIIE protein - Bacillus subtilis >gp
 :gp|Z99112|BSUB0009_150 Bacillus subtilis complete genome (s
 ection 9 of 21): from 1598421 to 1807200. NID: g2633902.
 atgattgatagcttttttaattatcttttggatgagtcgatatttaacttataatttta
 55 gtacttattgcaacaatttttataacatactctaagcaaatcctagaactcgacgtagt
 atcgggtgcaatagttttacaattagctttgtttatttatagcgcaattgtattttcatttt
 tcacataatatcacttctcaaagagagcctgtactgtcctttgtttataaagcttatgaa
 caaacacattttccaaattttgggggaggcttaataggtttttatttacttaaaactattt
 atacctctcatatctattgtaggtgtaataataattactatcctattactagcttcgagt

5 ttcatttttattacttaattttaagacatagagatgttacaaaaagtttattcgacaacctc
 aagtcacaaagtaatacatgcatctgagtcataaaaaacaaaaagagaacaaaaataagatt
 aaaaaagaagaaaaagcccaattaaaagaggcaaaaattgaacgaaaaaaacaaaaaaa
 tcacgtcagaataataatgtcattaaagatgtagtgattttccagagatttctcagtc
 10 gacgatattccaatatatggtcataatgagcaagaagataaaaagaccaaatactgctaac
 caacgtcaaaaacgtgttttggataatgaacaatttcaacaatcattaccaagtaccaa
 aatcaatcaataataataatcagccatctacaaccgctgaaaacaatcaacaacaaagt
 caggctgaaggctcaatatctgaagctggtgaagaagccaatatagagtatacggcgcca
 cctttatccttattaaaacagcctactaaacaaaaaactacttcaaaagctgaagtc
 15 cgtaaaggctcaggttttagaatctacactaaaaaactttggagttaatgctaaagtaaca
 caaattaaaatcggctcctgcagttacgcaatatgaaattcaaccagcgcaagggtgttaa
 gtagataaaatgaagctcaatctccataatgacattgcattagctttggctgcaagagtgta
 cgaatagaagcacctattccaggctcgtctgcggttaggaattgaggttcccaatgataaa
 atctcacttgtcactctaaaagaagttttagaagataagttcccatctaagtataaatta
 20 gaagtcggcattggttagagatatttctggtgatccaatatcaattcaattaaatgaatg
 cctcacttactcgttgctggttcaacaggaagcggtaaatcagtttgattaatggtatt
 ataacgagtatattactcaacacaaaaccgcacgaagttaaacttatgtaaatcgatcct
 aaaaaggtagagttaaatgtttacaatggtattcctcatttacttataccggttgtaaca
 aaccacataaaagcgtctcaagcttttagaaaaaattggttcagaaatggaacgtcgttat
 25 gatttgtttcaacattcatcgacacgaaatattgaaggatataaccaatatatacgcaaa
 cagaatgaagaacttgatgaaaaacaacctgagttaccgtatatcgtcgtaatagtggat
 gaattggctgatttaatatggttgaggtgaagaagtagaaaatgctatccaacgtatt
 actcaaatggctagagcagcggttatacacttaattgtagctactcaagaccttccggt
 gatgttattactggtatttataaaaaataacattccatcaagaattgcgttcgctgtaagt
 30 tctcaactgactctagaacaataattggtgctggtggagctgaaaagctacttggtaaa
 gtagatatgctatatgttggtgaacggagaatctactacaaccggaattcaagggtgctttt
 ttaagtgatcaagaagtgaagatgttggttaattatgttgtagagcaacagaaagcaaat
 tatgttaaagaatggaaccagatgcacctgtagataaatcagaaatgaagagtgaggat
 gctttatatgatgaagcttatttatttgaatagaaaagcaaaaagctagtacttcttta
 35 ttacaacgacaattttagaatcggttataatcgagcttcaaggctcatggatgatttgga
 cgtaaccaagttattggtccacaaaaaggaagtaaacctagacaaatattagttgattta
 gaaatgacgaggtgtaa

Sequence 1742

35 MIDSFFNYLFGMSRYLTYILVLIATIFITYSKQIPRTRRSIGAIVLQLALLFIAQLYFHF
 SHNITSQREPVLSEFVKAYEQTHFPNFGGLIGFYLLKLFIPILISIVGVIIITILLASS
 FILLNLNRHRDVTKSLFDNLKSSSNHASESIKQKREQNKIKKEEKAQLKEAKIERKKQKK
 SRQNNNVIKDVSDFPEISQSDDIPIYGHNEQEDKRPNTANQRQKRVLDNEQFQSLPSTK
 40 NQSIINNQPSTTKAENNNQSQSQAEGSISEAGEEANIETVPLSLKQPTKQKTTSKAEVQ
 RKQVLESTLKNFNGVNAKVTVIKIGPAVTQYEIQPAQGVKVSIVNLHNDIALALAAKDV
 RIEAPIGRSAVGIEVPNDKISLVTLEKVEDKFPISKYKLEVGIGRDISGDPISIQNLNEM
 PHLLVAGSTGSGKSVCINGIITSILLNTPHEVKMLLIDPKMVELNVYNGIPHLIPVVT
 NPHKASQALEKIVSEMERRYDLFQHSSTRNIEGYNQYIRKQNEELDEKQPELPYIVVIVD
 45 ELADLMMVAGKEVENAIQRITQMARAGIHLIVATQRPVSVDVITGIKNNIPSRIFAFAVS
 SQTDSRTIIGAGGAEKLLGKGDMLYVNGESTTTTRIQAFLSDQEVQDVVNYVVEQQKAN
 YVKEMEPDAPVDKSEMKSEDALYDEAYLFVIEKQKASTSLRQRFRIQYNRASRLMDLLE
 RNQVIGPQKGSKPRQILVDLENDEV*

Sequence 1743

50 Contig_0648_pos_4242_4955,
 putative peptide of unknown function
 atgtcggaatgagtgcaatctatagagtaaaacaatacatttttaaatttaatacaagat
 ggtgaactaaccatggaagtaaatcactagtaatttgcaattgagagcatttaaat
 55 gttaaaacagatgatgtttatgatggtatagatgagttgattactgaacaagtagtaacg
 gataattttgaagaggggactagcgtaaaagtaaaagcccccttctattaccggttaaat
 aaaattatttagtatagggactatgattaaagaagcgggttatgaagcaggaacagaatat
 ctgaatcttgacgagcaacctgcaactatttttagatgctgaacatttaggtatagaaca
 aaagaacctataacaattattgagagactaaggactgctaatcataagcctgtcgatatat
 tgttttagacaaaatagcaaaaacttatctaacttgtacagattatcaacagagtagtggt

tcaatgtagaagctattaaagcatctacaaatcatcaaatcatgcatgcagaaatggat
ttagaagcaattagttacgaaccccatatctctgaagtgttaatgttcacctcacgaa
gggcttatgttacttaaagtagtacattatgacgaaaagcatcaaccaattttgtattct
ttaaattatattaagagtagtttagttaattcactattactaaaagtgaataa

5

Sequence 1744

MSEMSAIYRVKQYILNLIKDGELTNGSKLPSNLSIARALNVKTDDVYDGIDELITEQVVT
DNFEEGTSVKVPFFYYPLNKIISIGTMIKEAGYEAGTEYLNLDQPATILDAEHLGIET
KEPITIIERLRTANHKPVVYCLDKIAKTYLTCTDYQQSSGSMLEAIKASTNHQIMHAEMD
LEAISYEPHISEVLNASPHEGLMLLKVVHYDEKHQPILYSLNLIKSSLVKFTITKSE*

10

Sequence 1745

Contig_0648_pos_5421_6257,

is similar to (with p-value 2.0e-28)

15

>gp:gp|AF082738|AF082738_1 Streptococcus pyogenes phosphatidylglycerophosphate synthase (pgsA) and ABC transporter ATP-binding protein (stpA) genes, complete cds; and unknown genes .
NID: g3426363.

20

atggaagataataaagcacagtagttcatttcttcagttaatgaattatattgtttaacaa
gaaccctatcgatatatagcgacaggtcaattagaacaaattccacaagtgacttctgaa
agtctatacgatacatatctatccatgggtacaaaatgatgattgtgccatatattgttga
ggaaatattaacaaagagggaagtaacgcaactaattctagataaagtttgcaattaagcct
ttctatttagaaaataaagaaagtactgaaatcacaccttcttttgatcaaccgcaatat
ataattgaaaaagacgatgttgaccaagctaaattgaatttgggatatcgctttccatct
tattatgggaaaagtaattactatgcatttatagttatgaatatgatgtttggaggagat
25 ccttcctcagtagtatttaaatgaagtcagagaaaagcaaagtttggcatactctatacat
tcacaaattgatggtaaaaacggattttttatttgttttaagtgggtgtttctgctgagaaa
tatgagcaagcaaaagatactgtcatcaaagagtttgataagataaaaaatggagatttt
gatttctaataaaaattgaattagctaaaaaaatcattatttcccatagacacgaagcatca
30 gatagacctaaaagtataattgaaatactacataatcaattattattaaaccgacagcaa
actgatcaagattttataaatgcagttaatcaagtgacgaaaaaagatgttattaaattg
gcaaatgaagctgttctagatacaatttatgtactaacgaaaggagaccaacactga

35

Sequence 1746

MEDNKAQYSFLQLMNYMFKQEPYRYIATGQLEQIPQVTSESLEYDYLTMVQNDDCAIYVV
GNINKEEVTLILDKFAIKPFYLENKESTEITPSFDQPYIIKDDVDQAKNLNGYRFPS
YYGKSNNYAFIVLMMFGDPSSVLFNEVREKQSLAYSISHSQIDGKNGFLFVLSGVSAEK
YEQAKDTVIKEFDKIKNGDFDSNKIELAKKIIISHRHEASDRPKSIIIEILHNQLLLNRQQ
TDQDFINAVNQVTKKDVIKLANEAVLDTIYVLTGDKQH*

40

Sequence 1747

Contig_0648_pos_6689_7012,

putative peptide of unknown function

45

atgtaccaggaacaaccaggatataaattaatgtttaatactttaagggctatgtattcc
aagcaccgatacgggtggatatcgctggttagtggtgaaagcatttatgaaataacaaaa
gatgatttatatctatgctatgagacattttatcatccctctaataatgggtgttgttgtg
gtaggcgatgttagtcctcaatcgataattaaactttagtagaaaagcatgaaaatcaaaga
aataaaacttatcaaccacgtattgaacgtgcgcaaattgatgagcctagagagataaat
caccggtttgtttctgagaaaaatga

50

Sequence 1748

MYQEOPGYKLMFNTLRAMYSKHPIRVDIAGSVESIYEITKDDLILCYETFYHPSNMVLFV
VGDVSPQSIKLVKHEHQNRNKTYQPRIERAQIDEPREINHGLFLRK*

55

Sequence 1749

Contig_0648_pos_7033_0,

putative peptide of unknown function

atgctaggttttaaaaaatgaaccattagatgaaagtgaactaaattttgttcaaagagat
ttggaaatgacattttttctacgaattgggtttttggagaggaaacggaatttttatcaacaa

cttttaataaagatttaataagatgaaacattcggttatcaatttgattggaaccgagc
 tacagtttttcaattattactagtgaacacacagcctgatctatttaacaatttaata
 atggatgaattaagaaaatataaaggaaaccttaagatcaagaagcatttgattgttg
 5 aaaaagcaatttattggagaattcatatcaagtttaattctccagaatatattgcta
 caatatgcaaaactctatttcgagggagtgagtgatttgatatgcttgatatcgtagaa
 aatattacggttagagagtgtaaataaaacttccgaattattcttgaaacttgaccaactt
 gttgatagtcggttgagatggaaaataga

Sequence 1750

10 MLGFKNEPLDESATKQVQDLEMTFFYELVFGEETEFYQQLLNKDLIDETFGYQFVLEPS
 YSFSIITSATQPPDLFKQLIMDELRLKYKGNLKDQEAFLKKQFIGEFISLNSPEYIAN
 QYAKLYFEGVSVFMDLDIVENITLESVNETSELFLNFDQLVDSRLEMENR

Sequence 1751

15 Contig_0649_pos_271_588,
 putative peptide of unknown function
 atgaatattttgatggtttataaaaaatgaatgatattgccataaacggacttatgttgctc
 atttctcacgcaattatgatattagaagctatctatttttatcctcggttttaaaatct
 aaattggctggattgatgagtttcatatgggtgacgatcaatgacgtaatagattacata
 20 taocgacaatatccctactatgattttatcgccaaacatttaattgaagtaggggtattg
 gcttatagtcctcactatcatttcgtatattttatttttaaaattacaaaagtgggtgaaa
 gttaaaacatttgattaa

Sequence 1752

25 MNILMFIKMNDIAINGMLLLISHAIMILEAIYFYPFKISKLAGLMSFIWVTINDVIDYI
 YGQYPYDFIAKHLIEVGVLAYSILTII SYILFLKLQKWLKVKTFD*

Sequence 1753

30 Contig_0649_pos_1920_2990,
 is similar to (with p-value 0.0e+00)
 >gp:gp|L38424|BACJOJC_6 Bacillus subtilis dihydropicolinate
 reductase (jojE) gene, complete cds; poly(A) polymerase (joj
 I) gene, complete cds; biotin acetyl-CoA-carboxylase ligase
 (birA) gene, complete cds; jojC, jojD, jojF, jojG, jojH gene
 35 s, complete cds's. NID: g755600.
 atggcagaacgtggccatgaagttcactttattacctcaaacataccctttagaatacgc
 aaacctttacctaacaatgacgttccaccaagttgaagtcaatcaatatgccgtattccaa
 tatccaccatacgaattacattaagtaaaaaattctgacgttatacaagaatgat
 ttagacatatgtcatatgcattatgctgtacctcacgctgtatgtggtatattagcgaaa
 40 caaatgtcaggtgaaaaacgtcaaaattatgacaacactacatggcactgatataactgtg
 ttaggttatgaccatactttacaaaacgcgataaaaatttggcatagaacaaagtgatatt
 gtaacaagtgttagccattctctagcacagcaaaacttatgaaattatcaatactaaaaag
 gaaatcatccctatatataattttataagggaaaatgaattcccaactcggcataatgaa
 gaattaaaagattgttatggtatttcacctgaagaaaaggtattgatacatgtttcta
 45 ttcagaaaagttaaaacgtattgatacagtgattgagacatttgcaaaagttcatgagagt
 ataccatccaagttgatacttttaggagatggtccagaattaatcgatatgcgacataaa
 gcacgagaacttgatgttgaaacacacgtactcttttaggcaaaacaaatgacgtaagc
 gcattctaccaactatctgatttagtactactcttaagtgagaaagaaagtgttgatta
 actctcttagaagcaatgaaaacaggcgtcttacctatagggagtcagcaggtggtatt
 50 aaagaggtcatcagacatgaagaaactggatttatagtagatataggggatagtacacaa
 gctgcaaaatagctattaaacttttatcaaatccagagttatatcaaaaatgcaatca
 caaatgctgaaagatattgaagcaagatttagttcagatttaattactgaccaatatgaa
 aactattatcgaaagatgctagaacaaggtgagaacaacaatgagtcatga

55 Sequence 1754

MAERGHEVHFITSNIPFRIRKPLPNMTFHQVEVNQYAVFQYPPYDITLSTKISDVIQEYD
 LDILHMHYAVPHAVCGILAKQMSGKNVKIMTTLHGTDTITVLGYDHTLQNAIKFGIEQSDI
 VTSVSHSLAQQTYEIINTKKEIPIYNFIRENEFPTRHNEELKDCYGISPEEKVLIHVS
 FRKVKRIDTVIETFAKVHESIPSKLILLGDGPELIDMRHKARELDVETHVFLGKQNDVS

AFYQLSDLVLLSEKESFGLTLLEAMKTGVLPISGSHAGGIKEVIRHEETGFIVDIGDSTQ
AAKYAIKLLSNPELYQKMQSMLKDIEARFSSDLITDQYENYYRKMLEQGENNNES*

Sequence 1755

- 5 Contig_0649_pos_3166_4182,
is similar to (with p-value 1.0e-51)
>sp:sp|P42977|PAPS_BACSU POLY(A) POLYMERASE (EC 2.7.7.19) (P
AP). >gp:gp|L47709|BACYPIA_15 Bacillus subtilis (clone YAC15
-6B) ypiABF genes, qcrABC genes, ypjABCDEFGH genes, birA ge
10 ne, panBCD genes, dinG gene, ypmB gene, aspB gene, asnS gene
, dnaD gene, nth gene and ypoC gene, complete cds's. NID: gl
146223. >gp:gp|Z99115|BSUB0012_185 Bacillus subtilis complet
e genome (section 12 of 21): from 2195541 to 2409220. NID: g
2634478.
- 15 gtggggaaagaacacggtacgatcaacgtggtctttcaaaatgacaattatgaaattact
acattcagatctgaagatgaatacatcgatcatcgtaggccaagtgaagtgtattttgta
agagacttatatcaagatgttcaacgtagagattttacaatgaatgctatagcaatggat
ttaaattaccggttggtatgattatttttaattggtcaacaagatataaacaatcgagtaatt
cgtactgttggtgtaccaagtgaaggttttcagaagacgcgcttcgtatcattagagga
20 ttacgttttcaatcacaacttaattttcaaatgattcagacacattacatgcaatgtct
tctcagatttcagatatacaatatttatccgttgaaacgtgtagtagtagagcttaaaaaa
cttatcatgggaaacaatgttaaaacaaagttttgaagtcattgcaaaacatgaaagcattt
aattatatacctttttcaaatcatttgagatgtctcatcttcatatagatgagcccatc
acatttgaacttttgattgcaatcttaattcgctccaacaaccaaagatatataaattaagc
25 accttgaaatcagcaatcaagaaaagcaactatcaaaaaatgggttacactcatccaa
acattgcctaagatacagtcacaagcaatctttaataacattagtagtatgattacaattta
aatgatattgaaatcctattatcattacatcatttgcttaaaaaaatgggcttacaaca
gccaatcatttaatcattaatgaaataagtattcgcgaaagcaaatgaaaaattacctatc
cattgcagaaaagaattggcaataaatggttaaagatatactcaatcatacgaataaaaaat
30 tcaggaccatggctaaaagatacacttagagaaatagaaatcgagtcattatcaaatcaa
atagtcaacactaaagaagaatattagaatgggtggatgcacatgtcaaaatatag

Sequence 1756

- 35 VGKEHGTINVVVFQNDNYEITTFRSEDEYIDHRRPSEVYFVRDLYQDVQRRDFTMNAIAMD
LNYRLYDYFNGQQDINNVRVIRTVGVPSERFSEDALRIIRGLRFQSQLNFQIDSDLHAMS
SQISDIQYLSVERVVVELKKLIMGNVVKQSFEVMQNMKAFNYIPFFKSFEMSHLHIDEPI
TFELWIAILIVQQPKDIQLSTLTKISNQEKATIKKWTLIQTLPKIQSKQSLITLVYDYNL
NDIEIILSLHLLKQNGLTANHLIINEISIREANEKLPICRKLALINGKDIILNHTNKN
SGPWLKDTLREIEIAVISNQIVNTKEEILEWVDAHVKI*

40

Sequence 1757

- Contig_0649_pos_4199_5140,
is similar to (with p-value 2.0e-47)
>sp:sp|P42975|BIRA_BACSU BIRA BIFUNCTIONAL PROTEIN (BIOTIN O
45 PERON REPRESSOR) (BIOTIN--[ACETYL- COA-CARBOXYLASE] SYNTHETA
SE) (EC 6.3.4.15) (BIOTIN--PROTEIN LIGASE). >gp:gp|L47709|BA
CYPIA_16 Bacillus subtilis (clone YAC15-6B) ypiABF genes, qc
rABC genes, ypjABCDEFGH genes, birA gene, panBCD genes, din
G gene, ypmB gene, aspB gene, asnS gene, dnaD gene, nth gene
50 and ypoC gene, complete cds's. NID: gl146223. >gp:gp|Z99115
|BSUB0012_184 Bacillus subtilis complete genome (section 12
of 21): from 2195541 to 2409220. NID: g2634478..
- atgttgatgaacatcaatcaaattacatcagggcaatatattgccgatcaactcaat
atttctagagcaggtgtcaaaaaagttattgacctattaaaagaagatggttgatc
aagtcaataaaccacaaaggccatcaactgaattcattacctgatcagtggtatagcgg
55 attgtaaaacctattctcgatgaacttggcctttttaatcatctagaagttatcacact
gtagattcaacacaattaaaagcaaaagagagcactcgttggaataaaagatacttttta
atgttgagcgatgaacaaaccgaaggtagaggttagattcaatcgtaattgggaatcatct
aaaggaagggttatggatgtcactagtgtcaagacctgacgtaccttttctatgata

cctaaatttaattttattttattgcttttaggtattagagatgctattcaacaattttcgaac
 gaacgtgtaacaattaaatggccaaatgatatatatttggaataaaaaaatttgcgga
 tttttaactgaaatggttgcaaatatgatgaaatagaagcaataatttggttataggt
 ataaatatgaatcatgttgaaagtgattttgacgaggatattaaagataaagcaacaagt
 5 atacgcatgcattccgatagatataattaatagatatatacttttttaactgcattattaact
 caaattatacatcgctttgatcaatttttacatcaaacttttgagtcaattcgagaagaa
 tatattcacgctacaaatatatggcatcgtaacttaaattcactgaaaataatcatcaa
 tttttgggggaagccatagatattgattcagatggattccttattgttaaagatgaaaaa
 ggtcaattacatcgacttatgagtgagatagatttataa

10

Sequence 1758

MLYEHQSNYISGQYIADQLNISRAGVKKVIDLLKEDGCDIKSINHKGHQLNSLPDQWYSG
 IVKPIDELGLFNHLEVYHTVDSTQLKAKRALVGKNDTFLILSDEQTEGRGRFNRNWESS
 KGKGLWMSLVLRPDVPFSMIPKFNLFIALGIRDAIQQFSNERVTIKWPNDIYIGNKKICG
 15 FLTEMVANYDEIEAIIICGIGINMNHVESDFDEDIKDKATSIRMHSDSIINRYTFLTALLT
 QIIHRFDQFLHQT FESIREEYIHATNIWHRQLKFTENNHQFLGEAIDIDS DGLIVKDEK
 QQLHRLMSADIDL*

Sequence 1759

20 Contig_0649_pos_5291_0,
 is similar to (with p-value 3.0e-66)
 >sp:sp|P54394|DING_BACSU PROBABLE ATP-DEPENDENT HELICASE DIN
 G HOMOLOG. >gp:gp|L47709|BACYPIA_20 Bacillus subtilis (clone
 YAC15-6B) ypiABF genes, qcrABC genes, ypiABCDEFghi genes, b
 25 ira gene, panBCD genes, dinG gene, ypmB gene, aspB gene, asn
 S gene, dnaD gene, nth gene and ypoC gene, complete cds's. N
 ID: g1146223. >gp:gp|Z99115|BSUB0012_180 Bacillus subtilis c
 omplete genome (section 12 of 21): from 2195541 to 2409220.
 NID: g2634478.

30 atgatacgtaccgatttagaaattccaccgtttatccaagccttaacttctatagaagaa
 gaaatgtagtacaggcaccttattttaatgaggtcgagatgacatttatcaacttatc
 aaagactgtgtttttgtgcacataatatctcatttgatcttaattttataaaaaaagct
 tttgaaaaatgtaatatattcaatttaaacctaaaagagtaattggatactttagaattggtt
 aaaatcgcatctcctacagacaaaagctaccagttaagtgcattggctgaatctcatcat
 35 ataccattaaataatgcacatagagcagatgaagatgcaacaacaactgctaaattgatg
 attaaagcttttgagaagttcgagcaattgcatttagatacacaacaaactgtactat
 ttaagtaaaaatctcaagtatgatctttatcatattttgtttgaaatggttagaaattat
 caaactaaaccgcctaacaaccaattttgaacaattttgaacaaattatttaccgtaaacaa
 attgatttaaaaaaacggctgtcaattttgatggtaccttaaaagatttatataaaaaat
 40 gtcactcaatcattaaatcttacatatcgccacaacagttataacctagccgaaattata
 cttgatcagttgatgcatagtgataaggcaatgattgaagctcctttgggtagtggaaag
 tctcttgcttacctgcttcgagcaacaatgtataatattgagaccggtcgatgtaagt
 atttcaacaaatacaaaattattacaaagtcagctattagagaaagacataccattactc
 aatgatgttttagattttaaaattaacgcgtcattaatcaaaagtaaaaatgattatata
 45 tctcttggtcttatcagccaaattcttaaaagcagatacaaaataattatgaagtaagtatt
 cttaaaaatgcagttacttatttggataactgaaacaaataactggggacatacaggaatta
 aatcttaaaaggtggacaaaaaatgtatgttgacaaaaaattgaaacatacgttccagtt
 cgtcatgatattcattattataattatataaaaaagaaatgctcaaaacatacaaaattggt
 atcactaatcatgcgcacttaattcattcagacagtgaaaacactatatatcaactattt
 50 gatgattgcatcatcgatgaagcacatagattgcctgactatgcgctaaatcaagttact
 aatgatttaaattattcagatgttaaataatcaattaggacttattggcaaaaatgaaaat
 gaaaaactacttaaaagcagtagacaaacttgagcaacaacgtattttagagaaactagat
 attgcacctatagacgtttttggactgaaaataaatatcaatgagttacatgatgttaaat
 gagcaactattcactacaaatttataatattattcaaacatcagacgtttatgatgatgac
 55 attcataagtatcattacgtttatgactttgaaacgggtgagattttaaaagatttacgt
 gcaatcatagataaattaaataaaacgatagaaatttttaacggaatgaatcacaaaaca
 atcaagctctgtacggaacaattattatacttacatgacaaatttaaacttata

Sequence 1760

MIRTDLEIPFFIQALTSIEEEMLVQAPYFNEVADDIYQLIKDCVFVAHNISFDLNFIIKKA
 FEKCNIIQFKPKRVMDTLELFKIAFPTDKSYQLSALAESHRIPLNNAHRADEDATTTAKLM
 IKAFEKFEQLHLDTOQKQLYYLSKNLKYDLYHILFEMVRNYQTKPPNNQFEQFEQIIYRKQ
 5 IDLKKPAVNFDTGLKDLKYNVTQSLNLTYPQQLYLAEIILDQLMHSBKAMIEAPLGSGK
 SLAYLLAATMYNIETGRHVMISTNTKLLQSOLLEKDIPLLNDVLDKINASLIKSKNDYI
 SLGLISQILKDDTNNYEVSLKMQLLIWITETNTGDIQELNLKGGQKMYVDQKIETYVPV
 RHDIIHYNYIKRNAQNIQIGITNHAHLIHSSENTIYQLFDDCIIDEAHRPLPDYALNQVT
 NDNLNSDVKYQLGLIGKNENEKLLKAVDKLEQQRILEKLDIAPIDVFLKININELHDLN
 10 EQLFTTIYNIQTSDDYDDDIHKYHYVYDFETGEILKDLRAIDKLNKTIEIFNGMNHKT
 IKSVRKQLLYLHDKFKLI

Sequence 1761

Contig_0650_pos_3702_4013,
putative peptide of unknown function

15 gtgtataatggtataaatgttatgtatataattatatggaggcattatatgaagagtatg
 aagcaaatcgctgatgaattaaacgtaacaaagatgactgtttataataatgctaagaaa
 gcaaatgtgaaatttcaaaaaattgaaaatgtaaattttatcttccgaagatgaagtt
 atagtagctaatagaataaaaaaaaaatcaaaaataaaactgattacttcgataatgaaaaa
 20 aaggtagaaccaaacccaacaatgataatcttgtaaaaatgaaacgattaaacatttat
 ataaccaattag

Sequence 1762

VYNGINVMYIIWRHYMKSMKQIADELNVTKMTVYNNAKKANVKFQKIENVNYLSSEDEV
IVANRIKKNQKTDYFDNEKKVETKPNNDNLVKMKRLNIYITN*

25

Sequence 1763

Contig_0650_pos_8636_9154,
is similar to (with p-value 4.0e-82)

30 >gp:gp|U50077|SAU50077_1 Staphylococcus aureus multidrug res
 istance plasmid pKH8 replication protein (rep) gene, qacC' g
 ene, and multidrug resistance protein (qacC) gene, complete
 cds. NID: g1236637.

atgacgaaaagtgggaaacaacgcccattggagagaaaagaagatagataatgtaagttat
 gcagataacttgaaaattttaaaaataaaaaaggcttttaattgtaaaacaatgtggtaac
 35 gtcttagagttcaagccgactgatgaaggttatttgaagttacataagacatggtttgt
 aagtcgaaactctgccagtttgtaattggaggcgtgctatgaaaaatagttatcaagct
 caaaaagtgattgaagaagttgttaagaaaaacaaaagcgcgttggtatttttaaca
 ctttcaacgaaaaatgcatagatggggatactttagaacaaaagtttgaaacatttaacg
 aaagcatttgatagggttaagtagatataaaaaagtgaaagcaaaatcttggtgggttttg
 40 cgttcaacggaagtaacagtttaataaaaatgatggtagttataatcaacaattcgaacc
 tacacagtcgcaaaaaagtatgatagaagaattgattaa

Sequence 1764

45 MTKSGKQRFWREKKIDNVSYADILEILKIKKAFNVKQCGNVLEFKPTDEGYLKLHKTWFC
 KSKLCPVCNWRRAMKNSYQAQKVEEVVKEKPKARWFLTLSTKNAIDGDTLEQSLKHLT
 KAFDRLSRYKKVKQNLVGLRSTEVTVNKNDSYNQQFGTYTVAKKYDRRID*

Sequence 1765

Contig_0650_pos_9204_9533,
putative peptide of unknown function

50 atgttaaaagaagatatgaagttgccaaaatcttatattttgaaattgcttccaactgg
 aagaaaattggtatttcaaattgccaaacaagcatatgaatatgcattacaagttaatacaa
 ctaaaaaattacgaaacacatttctaataacgacagacaacaatcgtggaagacaaaat
 caatttttatccaagaaaagacacctaataatggcttcaaaaataggacgatcaagaagaa
 55 aataaaagaaataaatgatgacactctcgaagaagatcgacaagcatttcttgaaaagfta
 aatcaaaaagtggagggaagataactaa

Sequence 1766

MLKEDMKLPKSYIFEIASNWKIGISNAKQAYEYALQVNQPKNYETHSNDKRQNNRGRQN

QFLSKEKTPKWLQNRDDQEENKEINDDTLEEDRQAFLEKLNQKWKEEDN*.

Sequence 1767

Contig_0650_pos_9554_10453,
 5 is similar to (with p-value 6.0e-63)
 >sp:sp|P06567|DNAI_BACSU PRIMOSOMAL PROTEIN DNAI. >pir:pir|B
 24720|IQBS44 dnaA protein homolog, 44K - Bacillus subtilis >
 gp:gp|AF008220|AF008220_192 Bacillus subtilis rrnB-dnaB geno
 mic region. NID: g2293135. >gp:gp|X04963|BSDNAB_1 Bacillus s
 10 ubtilis dnaB gene for initiation of chromosomal replication.
 NID: g39880. >gp:gp|Z99118|BSUB0015_163 Bacillus subtilis c
 omplete genome (section 15 of 21): from 2795131 to 3013540.
 NID: g2635200. >gp:gp|Z75208|BSZ75208_2 B.subtilis genomic s
 equence 89009bp. NID: g1769994.
 15 atgggcgattctcaaaatctagataaacgtatacaaaaaataaaacaaatgtaatcaat
 gatactgacgttaaacattttcttgagaaaaatcgtagaatataactaatgagatgata
 gacgaagatttaaatgttcttcaagagtataaagatcaacaaaaagtttatgatggacat
 cgctatgatgattgtccgaattttgtaaaaggacatgttctgaaactatatattgaaaat
 gaaagaatcaaaattagatatctaccttgcccggtgtaaaattaaacatgatgaggaacga
 20 tttgattcacaaacttattacatctcaccatattgcaaagagatacacttcatgcaaagctc
 aaagatatattatgaataatcgagagagacttgatgtagcaatggcagctgatcaaatc
 tgtacagcaattactaacgatgaaaaagtaaggggttatatttatatggtccttttggt
 acaggaaaatcattcatattgggtgctattgcaaatcaacttaaatcgcaaaagatttca
 tcaacaattgtatattaccagaattttattcgcaactttaaaaggtggctttaaaagacggt
 25 agttttgagaaaaaattacaacgtgtgacgagaagctaataattttgatgttagatgatatt
 ggcgcagaagaagtcacaccgtgggtaagagatgaagtgattggctctttattacattat
 agaatggtacatgaacttctacatttttttagttctaaactttaattatagttagcttgag
 catcatctttcaataactagagatggcactgaaaagactaaagcagcacgaattattgaa
 agaattaagactttatcgacaccttattatttgactggtaaaaaatttagaacaattga
 30

Sequence 1768

MGDSQNLDKRIQKIKQNVINDTDVKHFLEKNRSNITNEMIDEDLNVLQEYKDQKQVYDGH
 RYDDCPNFVKGHVPELYIENERIKIRYLPCPKIKHDEERFDSQLITSHMQRDTLHAKL
 35 KDIYMNNRERLDVMAADQICTAITNDEKVKGLYLYGPFGTGKSFILGAIANQLKSQKIS
 STIVYLPFIRTLKGGFKDGSFEKKLQVRANILMLDDIGAEVTPWVRDEVIGPLLHY
 RMVHELPTFFSSNFNYSELEHLSITRDGTEKKAARIIERIKTLSTPPYYLTGKNFRNN*

Sequence 1769

Contig_0650_pos_10806_12743,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P18255|SYT1_BACSU THREONYL-TRNA SYNTHETASE 1 (EC 6.1.
 1.3) (THREONINE--TRNA LIGASE) (THRRS). >pir:pir|B37770|YSBST
 45 1 threonine--trna ligase (EC 6.1.1.3) 1 - Bacillus subtilis
 >gp:gp|AF008220|AF008220_195 Bacillus subtilis rrnB-dnaB gen
 omic region. NID: g2293135. >gp:gp|M36594|BACTRNASB_1 B.subt
 ilis threonyl-trna synthetase (thrSv) gene, complete cds. NI
 D: g143765. >gp:gp|Z99118|BSUB0015_160 Bacillus subtilis com
 50 plete genome (section 15 of 21): from 2795131 to 3013540. NI
 D: g2635200. >gp:gp|Z75208|BSZ75208_5 B.subtilis genomic seq
 uence 89009bp. NID: g1769994.
 atgaatcaaatattcaatttccagatggtaatacaaaaagaatttgataaagggact
 actacagaagacatcgctcaatcaattagtcaggattaagaaaaaagcagttgcggga
 55 aaattcaatgggtcaactttagatttaaacacgccctttagaacaagatggagctattgaa
 attattactcctgggagtgagaagcggttagaagtacttcgtcattcaacagctcattha
 atggcacaagcattaaaacgtttatacggagacgttaaatttgaggttgacctgtaata
 gaaggcggattctattatgattttgatatggatgataaggttcatcggatgattttgat
 aaaattgagaaaaaatgaacaaattgtgaacgaaaatcataaaattgtaagagaagta

gttagtaaagaaaaagcaaaagacttcttcaaggatgacccttataaattagaacttatt
 gatgcaattcctgaagatgagagtgtaacactttataactcaaggtgaatttactgattta
 tgtcgagggtgtacacgtaccttctacttctaaaaattaaagagttcaaaactattatctaca
 gctgggtgcttattggcgtggaaatagtataataaaaatgttacaacgaatttatggtaca
 5 gcattctttgacaaaaaagatttgaaagcacatctaaaaatgttggaagaacgtcgtag
 cgtgatcatcgtaaaattcgtaaaagatttagaattgtttacaaacaatcaactcgttggg
 gctgggttaccattatggttaccaaatggtgctacaatacgtagggaatagaacgttat
 attgtcgataaagaagtaagtatgggatacgtatcatgtttacacaccagttatagccaat
 gttgatttatataaaacatctggtcactgggatcattatcaagaagatatgttcccagca
 10 atgaagttagatgaagacgaagcaatggtcttaagaccaatgaactgtccacatcatatg
 atgatttataaaaacaaacctcattcttctcggaattacctatacgtattgctgaattg
 ggtactatgcatcgtttacgaagcaagtggtgcagatcagggtttacaacgtgttcgagga
 atgacattgaatgatttcccatatttctggttagacctgatcaaattaagaagaattttaa
 cgtgtagttaatatgattcaagatgtgtacaaagattttgggtttgaagattatcgcttc
 15 agattgagttatagagatcctgaagataagcataagtaactttgatgatgatgaaatgtgg
 gaaaaagctgaatccatgcttaagaagcatcagatgaattaggtttaacttatgaagaa
 gctattggtgaggcagcattctatggacctaaagttagatgttcaagtaaaaaacagctatg
 ggaaaaagaagaactctatcaacagcacaaactgattttcttttaccagaacgttttgac
 ttaacgtacattggtcaagatggagaacaacatcgctcgttagttatacaccgtggtgta
 20 gtttctactatggaaacgttttgttgcaatttttaacagaagaaacaaaaggtgcatttcca
 acttggttggcgcctatgcaagttgaaattatcctgtaaatatagatttacattatgat
 tatgcaagactttttacaagatgaactaaaatcgcaaggtgtccgcgttgaaattgatgac
 cgtaatgaaaaaatgggatataaaattcgtgaagctcaaatgaaaaaaatcaccttatcag
 attggtgtaggtgaccaagaagtagagaatcaagaagtaaatgaagaaaatatggttct
 25 gaaaaacaagaatcagttgaaaaagatgaatttatttgggaatgttattgatgaaatccgt
 ttgaaaaagcatagataa

Sequence 1770

MNQINIQFPDGNTEKFDKGTTEEDIAQSISPLRKKAVAGKFNGQLVDLTPLEQDGAIE
 30 IITPGSEALEVLRHSTAHLMAQALKRLYGDVKFGVGPVIEGGFYDFDMDKVSDDFD
 KIEKTMKQIVNENHKIVREVVSKEKAKDFFKDDPYKLELIDAIPEDESVTLYTQGEFTDL
 CRGVHVPSTSKIKEFKLLSTAGAYWRGNSDNKMLQRIYGTAFDCKDLKAHLKMLEERRE
 RDHRKIGKDLELFTNNQLVGAGLPLWLPNGATIRREIERIYIVDKEVSMGYDHVYTPVLAN
 VDLYKTSGHWDHYQEDMFAMKLDEDEAMVLRPMNCPHMMIYKNKPHSYRELPIRIAE
 35 GTMHRYEASGAVSGLQVRGRTLNDSHIFVRPDQIKEEFKRVVNMIQDVYKDFGFEDYRF
 RLSYRDPEDKHXYFDDDEMEKAESMLKEASDELGLTYEEAIGEAIFYGPKLDVQVKTAM
 GKEETLSTAQLDFLLPERFDLTYIGQDGEQHRPVVIHRGVVSTMERFVAFLEETKGAFF
 TWLAPMQVEIIPVNIDLHYDYARLLQDELKSQGVVVEIDDRNEKMGYKIREAQMKKIPYQ
 40 IVVGQDEVENQEVNVRKYGSEKQESVEKDEFIWNVIDEIRLKKHR*

Sequence 1771

Contig_0650_pos_5141_4719,
 putative peptide of unknown function

gtgttgtttatgtataaatcaattttattagcagccgatgggtcagaaaatagtttacgt
 45 tcagcacaggaagttttgaactttatagatgaaaatactatagttactttaattacagtt
 gtaaatgttgaagaatcgaaaacagatgttttacatggttaacaaggacatagtttaaca
 aatgaaagagaagacaaattatccagtataactgaactattttagaacataatgtaaat
 tatgaagtaaaaaattgcacatggtcttctcgcagaaacagtggtttcagttgctaatagt
 ggtaaatatcaagcaattgttttagggctctcggtggtctaaatagtttacaagaaatggt
 50 ttgggtagcgtcagtcacaaagtggttaaacgttcaaaaattcccgttatcattgtaaaa
 tag

Sequence 1772

VLFMYKSILLAADGSENSLRSAQEVNLFIDENTIVTLITVVNVVEESKTDVLHGKQGHSLT
 55 NEREDKLSSITELFVEHNVNVEVKIAHGLPAETVVSVANSKGKYQAIVLGSRGLNSLQEMV
 LGSVSHKVAKRSKIPVIIVK*

Sequence 1773

Contig_0650_pos_1503_1072,

putative peptide of unknown function

atgaagaatatggtaatcttgaataagcaaaaaaggatgatcagaatgaaaaaagcaata
tttagtattattatttcttcttatttttagttctaactgctactggatgtagtaatagttct
aaagaaaaaccaattaaaaaaagtgcattagaaattaatcctacaagtaaagctgttaat
5 attacagtaaataaaaaagaaaataacaaacctgaaaaaattgggaaagtgtatcgatat
aaaaataacaatgcaaaaagaaattactaacgcagcgtattaaaaagataactaaagataca
ttgatttggaaaggtgtagcaaacaaatcagataatgtaaaagatttattaggagaaagt
attccttatgaagttaaatataaaaaatgggatataaaaaaattcgagagaaaaattaaa
tatactgaataa

10

Sequence 1774

MKNMVILNKQKRMIRMKKAIFSIIISLILVLTATGCSNSSKEKPIKKSALINPTSKAVN
ITVNNKENNKPEKIGKVYRYKNNNAKEITNDGIKKDKDTLIWKGVANKYDNVYKDLLGES
15 ILYEVKYKNGDIKKFERKIKYTE*

Sequence 1775

Contig_0650_pos_627_256,

putative peptide of unknown function

gtgtatattcattataatcacttagatcctaagcatgccaatgatgaaacctggatctg
20 ttaaaattattacattttagatcaagtttaaggatcatcatccttttgaaatatcaacaggt
caaaagcgtcgttttaagtgttgcaacagcattaagttcaaaggcagagattattttacta
gatgaaccaacattcggcctagatagtcataatacatttcaacttattaagttatttcaa
gaacgcgttaatcaaggtcaaacattatcatgggtgacacatgatccagaaattattaaa
cgatatccaacaagacgattacgcgtggaagatggatgtcttaagaaatggaaggtgaa
25 cacattgtttga

Sequence 1776

VYIHYNHLDPKHANDETMDLLKLLHLDQVKDHPFEISTGQKRRLSVATALSSKAEIILL
DEPTFGLDSHNTFQLIKLFQERVNQGTIIMVTHDPEIIKRYPTRRLRVEDGCLKEMEGE
30 HIV*

Sequence 1777

Contig_0651_pos_1791_2105,

putative peptide of unknown function

atgatacgtctttccattcgactgagtttgacttcgcttcttttaaagctaattgttaat
35 tcttttctaattatcgatctttctatatcgtttaatgccaaagtagcggttatattcaaca
atgtattccttaccgacagcttcccattttatagatactgttatgataattccgagtaaa
gggacaatggacaagattaacatagttatcggttagacctactgcagcaagaatgattgga
aagacaaaagtacctatgatactaccagtagcacttacagattctacaaagccagtagct
40 tgtgaacgttaaatga

Sequence 1778

MIRLSIRLSLTSLLLKANVNSFLIIDLSISFNAKVALYSTMYSLPTASHFIDTVMIIPSK
45 GTMDKINIVFVRPTAARMIGKTKVPMILPVRLTDSTKPVACERK*

Sequence 1779

Contig_0651_pos_4670_0,

is similar to (with p-value 0.0e+00)

>gp:gp|AJ005646|SAU5646_1 Staphylococcus aureus sdrD gene. N
50 ID: g3550593.

atgaaaaagagaagacaaggaccaattaacaagagagtggtattttctatccaacaaggta
aacaagtactcgattaggaagttcacagtaggtacagcttcaatactcggtgggtgctacg
ttaatgtttggtgccgcagacaatgaggctaaagcggctgaagacaatcaattagaatca
gcttcaaaagaagaacagaaggtagtcgtgataatgaaagctcaaaacttaatacaagtc
55 gatttagacaacggatcacatagttctgagaaaacaacaatgtaaacatgcaactgaa
gtaaaaaaagttgaagcaccacgaagtgacgtatctaagcctaagctaatgaagca
gtagtgacgaacgagtcactaaacaaaaacaacagaagcaccacactgttaatgaggaa
tcaatagctgaaacaccccaaacctcaactacacaacaagattcgactgagaagaataat
ccatcttttaaaagataatttaaatcatcctcaacgacatctaagaaagttaaacagac

gaacattctactaagcaagctcaaagtgtctactaataaatcaaatttagacacaaatgac
tctccaactcaaagtgagaaaacttcatcacaagcaataacgacagtagacacaatcag
tcagcaccttctaacaattagattcaaaacccatcagaacaaaaagtatatataaaacaaaa
5 ttaaatgatgaacctactcaagatggtgaacacacgacaactaaattaaaaacaccttct
atttcaacagatagttcagtcattgataagcaagattacacacgaagtgtgttagctagt
ttaggtgttgattctaatagaacagaagcaattacaaatgcagtttagagataaatttagat
ttaaagctgcattctagagaacaaatcaatgaagcaatcattgctgaagcactaaaaaaa
gacttttctaaccttgattatggtgtcgatagccattagctctgaacacatctcaatca
10 aaaaattcaccacataagagtgaagtcacgcattgaatttaattaggttttagctgtgag
cctaattagtggtaaaaatgtgaatgataaagttaaaatcacaaacccatcgctttcactt
aataagagtaataatcacgctaataacgtaatatggccaacagtaacgaacaatttaatt
ttaaagcaaattatgaattagatgacagcataaaagagggagatacttttactattaaag
tatggtcagtatatttagaccgggtggtttagaacttctgcaataaaaaactcaactacgt
agtaaggtggtcattttagctaatggtgtatatgataaaaactacaaatcacgacgact
15 tatacatttactaactatggtgatcaatatcaaaatattacaggtagttttgatttaatt
gcgacgcctaagagggaaacagcaattaaggataatcagaattatcctatggaagtgcg
attgctaacgaagtagtcaaaaaagacttcattgtggattatggtataaaaaaggacaat
acaactacagcagcggtagcaaatgtggataatgtaataataaacataacgaagtgtt
tatctaaaccaaataaccaaatacctaataatgctaaatatttctcaacagtaaaaaat
20 ggtaaatttataccaggtgaagtgaagtttacgaagtgcgataccaatgcgatggtg
gatagcttcaatcctgattttaaattagtttcaatgtaaaagatgtgacaagtcaatttaca
cctaaagtaagtgcagatggtactagagttgatataattttgcctagaagtatggcaaat
ggtaaaaagtatattgtaactcaagcagtgagaccaacgggaactggaaatgtttatacc
gaatattggttaacaagagatggtactaccaatacaaatgatttttatcgtggaacgaag
25 tctacaacggtgacttatctcaatggttcttcaacagcacaggggataatcctacatat
agtctaggtgactatgtatggttagataaaaaataaaaacgggtgttcaagatgatgatgag
aaaggttttagcaggtgtttatgttactcttaagacagtaacaatagagaattacaacgt
gtaactactgatcaatctggacattatcaatttgataatttacaaaatggaacgtacaca
gtcgagtttgcgatttctgataattatagccatctccgcaaataattctacaaatgat
30 gcaatagattcagatggtgaacgtgatggtacacgtaaaagtagttgttgccaaaggaca
attaataatgctgataatgactgtagatactggttttatttaactcctaataacaat
gtcggagattatgtatgggaagatacaaaataaagatggtatccaagatgacaatgaaaag
ggaatttcaaattgtcaaagtgcgttaaaaaataaaaatggagataccattgggacaacg
acaacagattcaaattggttaaatatgaattcacaggttttagagaacggggattacacaata
35 gaatttgagacgcccgaaggctacacacgactaaacaaaactcgggaagtgcgaaggt
aaagattcaaattggtacgaaaacaacagtcacagtcacaagatgcagataataaaacaata
gactcaggtttctacaagccaatatataacttaggtgactatgtatgggaagatacaaat
aaagatggtattcaagacgacagtgaaaaagggatttctggtgttaaagtgcgttaaaaa
gataaaaatggaatgccattgggacaacgacaacagacgcaagtggtcattatcaattt
40 aaaggattagaaaatggaagctacacagttgagtttgagacaccatcaggttatcacccg
acaaaagcgaattcaggtcaagatataactgtagattccaacggtataacaacaacaggt
atcatttaacggagctgataatctcacaattgtagtggtttctacaaaacaccaaataat
agtgtcggagattatgtatgggaagatacaaaataaagatggtatccaagatgacaatgaa
aaaggaatttctggtgttaaagtaacgttaaaggatgaaaaaggaaatataattagcact
45 acaacaactgatgaaaatgggaagtatcaatttgataatttagatagtggttaattacatt
attcattttgagaaaccggaaggcatgactcaaaactacagcaaattctggaaatgatgat
gaaaaagatgctgatggggaagatgttcgtgtaacgattactgatcatgatgacttttagt
atagataatggttattttgacgatgattcagacagtgactcagacgcagatagtgattca
gactccgacagtgactcggacgcagacagcgattctgacgcagac

50 Sequence 1780

MKKRRQGPINKRVDFLSNKNVYKYSIRKFTVGTASILVGATLMFGAADNEAKAAEDNQLES
ASKEEQKGSRDNESSKLNQVDLDNGSHSSEKTTNVNNATEVKKVEAPTTSDVSKPKANEA
VVTNESTKPKTTEAPTVEESIAETPKTSTTQQDSTEKNPNPSLKDNLNSSSTTSKESKD
55 EHSTKQAQMSNKNLNDNSPTQSEKTSQANNDSTDNQSAQSKQLDSKPSEQKVYKTK
FNDEPTQDVEHTTTKLKTPSISTDSSVNDKQDYTRSAVASLGVDSENETEAITNAVRDNL
LKAASREQINEAIIAEALKKDFSNPDYGVDTPLALNTSQSKNSPHKSASPRMNLMSLAAE
PNSGKNVNDKVKITNPTLSLNKSNHANNVWPTSNQFNKANYELDDSIKEGDTFTIK
YGQYIRPGLLELPAIKTQLRSKDGSIIVANGVYDKTTNTTTYTFTNYVDQYQYQITGSFDLI

ATPKRETAIKDNQNPMEVTIANEVVKKDFIVDYGNNKDNNTTAAVANVDNVNNKHNEVV
 YLNQNNQNPKYAKYFSTVKNGKFIPGEVKVYEVTDTNAMVDSFNPDLNSSNVKDVTSQFT
 PKVSADGTRVDINFARSMANGKKYIVTQAVRPTGTGNVYTEYWLTRDGTNTNDFYRGTK
 STTVTYLNGSSTAQGDNPTYSLGDYVWLDKNKNGVQDDDEKGLAGVYVTLKDSNNRELQR
 5 VTTDQSGHYQFDNLQNGTYTVEFAIPDNYTPSPANNSTNDAIDS DGERDGT RKVVVAKGT
 INNADNMTVDTGfYLT PKYNVGDYVWEDTNKDG IQDDNEKGISNVKVT LKNKNGDTIGTT
 TTDSNGKYEFTGLENGDYTIEFETPEGYTPTKQNSGSDEGKDSNGTKTTVTVKADANKTI
 DSGFYKPIYNLGDYVWEDTNKDG IQDDSEKGISGVKVT LKDKNGNAIGTTTTDASGHYQF
 KGLENGSYTVEFETPSGYTPTKANSQDITVDSNGITTTGIINGADNLTIDSGFYKTPKY
 10 SVGDYVWEDTNKDG IQDDNEKGISGVKVT LKDEKGNIISTTTTDENGKYQFDNLD SGNYI
 IHFEKPEGMTQTANSNGNDEK DADGEDVRVTITDHDDFSIDNGYFDDSDSDSDADSDS
 DSDSDSDADSDSDAD

Sequence 1781

15 Contig_0651_pos_1050_601,
 is similar to (with p-value 1.0e-71)
 >gp:gp|U96619|SAU96619_2 Staphylococcus aureus NCTC 8325 Sec
 E (secE), NusG (nusG) and RplK (rplK) genes, complete cds. N
 ID: g2078375.
 20 atgaatatgactgaacaaat ttttagagttgtcataccagaagaggaagaaactcaagtt
 aaggatgggaaagctaaaaagattgtgaagaaaacatttcctggatatgtattagttgag
 ttaatcatgacagatgagtcgtggtatgtagttagaatactcctggagtaacaggattt
 gtcggatctgcaggtgcaggatcaaaacctaaccctctacttcctgaagaagtacgcttc
 attcctaagcaaatgggtcttaagagaaaaacaatagatggtgaactcgatggtgggaa
 25 caagttcgtatccaatcaggtccttttgctaatacaattggagaagtacaagagattgaa
 gcggataaattcaagcttactgtactgttgatgattggtcggtgaacacactgtagaa
 gttgaatttgaccaaattgaaaaattataa

Sequence 1782

30 MNMTEQIFRVVPIPEEEETQVKDGKAKKIVKKTFFPGYVLVELIMTDESWYVVRNTPGVTGF
 VGSAGAGSKPNPLLPEEVRFILKQMLKEKTIDVELDVGEQVRIQSGPFANQIGEVQEIE
 ADKFKLTVLVDMFGRETPVEVEFDQIEKL*

Sequence 1783

35 Contig_0652_pos_3965_4576,
 is similar to (with p-value 8.0e-25)
 >gp:gp|AL031124|SC1C2_14 Streptomyces coelicolor cosmid 1C2.
 NID: g3355667.
 atgagagtaaaaaaattgataagtcattactgtatctagaataacaagtacaagaaac
 40 gtaattattgtaataaagagagaaggggtataataatgacaaaatttaactttgatcaa
 gttcacagtgatattcagtttaaaattaaacattcttatggtgtcccaagttaaagggaaca
 ttttaagcaattcgatggttcaattagatggagatattaatgatttaacttcaactaaaagca
 acagctactattattccaagttcaattgacactcaaaatgaggacagagacaaccattta
 agatcaaacgattttctttggtacagaagacaacgataaaatgacattttgtaactaaagaa
 45 attaacgaaaatcaagtagttggagatttgacaattaaaggtgaaactcatgaagagaca
 tttgatggtgaatttaaggtgaagtaaaaaatccaatgaatggacaacaagtcactgggt
 tttatcggttagtggaacaattaacgcgaaaaaatatgggtattaattttaaccaagcttta
 gaaactggtggcgtgatgttaggtaaaaacgtaaaatttgaagcatcagcagaatttagc
 atcgacaattaa

Sequence 1784

50 MRVKKLISHLTVSRITSTRNVIIIVIKREGVIIMTKFNFDQVHSDIQFKIKHLMVSQVKGT
 FKQFDVQLDGDINDLTSKATATIIPSSIDTQNEDRDNHLRSNDFGTEEDNDKMTFVTKE
 INENQVVGDLTIKGETHEETFDVEFNGVSKNPMNGQQVTGFIVSGTINREKYGINFNQAL
 55 ETGGVMLGKNVKFEASAEFSIDN*

Sequence 1785

Contig_0652_pos_6095_7057,
 is similar to (with p-value 0.0e+00)

>gp:gp|U92974|LLU92974_13 Lactococcus lactis unknown gene, partial cds, and HisC (hisC), unknown, HisG (hisG), unknown, HisB (hisB), unknown, HisH (hish), HisA (hisA), HisF (hisF), HisIE (hisIE), unknown, unknown, LeuA (leuA), LeuB (leuB),
 5 LeuC (leuC), LeuD (leuD), unknown, IlvD (ilvD), IlvB (ilvB), IlvN, IlvC (ilvC), IlvA (ilvA), AldB (aldB) and aldR (aldR) genes, complete cds. NID: g2565137.
 atggattatagagtactactttattataaatatgtaactatagatgaccctgaaactttt
 gcagccgaacattttgaaattttgtaaggaacatcatttaaaaggaagaatactagtttca
 10 acggaaggcattaatggaacattatctggaacaaaagaagatactgataaatatagag
 catatgcatgcagatagtcggttttgctgatttaacttttaaaattgatgaagctgaaagt
 catgcgttttaaaaagatgcacgtgcgtccaagacgtgaaattgttgcaacttgacttagaa
 gaagatattaatccacgtgaaattaccggttaatactattctcctaagaatttaaaagcc
 gcactagaagatgaaaatactggttatattagatgctcgaaatgattatgaatacgtattta
 15 ggacatttccgtggagctattcgtcctgatataacacgattccgtgacttacctgaatgg
 gtgcgtaataataaagaacaactcgacggaaaaaatattgtccatatattgtacaggtggc
 attcgtttgtgaaaaattttctggttggttagtaaaagaaggatttgaaaacgtaggtcag
 ttgcattggtggtattgtctacatacggtaaaagaccctgaaactaaagggctatattgggat
 ggtaagatgtagtattgtgaacgtatttagtgcgatgtgaatcaaattgataaaaca
 20 gtcacgcggaagagcattttgatggtacaccttgtgaacgttatattaattgtgcaaac
 cctgaatgtaataaacaattcttgtttctgaagaaatgaagaaaaatatttaggtgca
 tgttcgtatgattgtgcaaacatgagcgcaatcgctacgttgcctcatcatattagc
 aatgaagaatggcaacgtcgtttaataatttcaaagatgtgcctgaacacacacatgca
 taa

25 Sequence 1786
 MDYRVLLYYKYVTIDDPETFAAEHLKFCKEHHLKGRILVSTEGINGTSLSGTKEDTDKYIE
 HMDHDSRFADLTFKIDEAESHAFKKMHVRPRREIVALDLEEDINPREITGKYISPKEFKA
 ALEDENTVILDARNDYEDLGHFERGAIRPDITFRDLPEWVRNNKEQLDGKNIVTYCTGG
 30 IRCEKFSGLVKEGFENVGQLHGGIATYGKDPETKGLYWDGKMYVFDERISVDVNQIDKT
 VIGKEHFDGTPCERYINCANPECNKQILVSEENEKYLGCASYDCAKHERNNRYVARHHIS
 NEEWQRRLLNNFKDVPETHA*

Sequence 1787
 35 Contig_0652_pos_8373_7876,
 putative peptide of unknown function
 gtgaatgtgcataaaatagattttatcaggcaacaaatttcaaatccaacgattttgttctg
 ttgcaaatgtattggcgctatttacaatactatttacttataaatgggcttatcaaaaca
 acgcatatcattgaacaaaatcttgcacgaatcttatttttgattttaggtttcgca
 40 ctactagtatttttgcacgagtttattcatcgatattttgttcattatattttctaaaggt
 gaaaaaccatctttaaaatatgataaaaacaaaattattgtacagttctctcagacttgt
 tttcatcggtggcaatttacaattatcatgatagcaccacttggtatcataagtgcgacc
 ttactagcacttattcataggaatatcaagaggaggttgatacctcttgctaaacca
 tcacccaaaagttaataatattgctcctaccacaaacctgacatgattataacatgtatatt
 45 ttatggccgacaagttga

Sequence 1788
 VNVHKIDLSGNKFQIQRFVLLQIVLALFTILFTYKWAYQTTHIEQNLVMNLIFGFVGFA
 VLVILHEFIHRILFIIFSKGEKPSLKYDKNKIIVQFSQTCFHRWQFTIIMIAPLVIISAT
 50 LLALIHTGISRGCCIPLAKPSPKVNNIAPTKPDMIITCIFLWPTS*

Sequence 1789
 Contig_0652_pos_5633_4836,
 putative peptide of unknown function
 55 gtgtttgaagggaatatttttaagttaatcccgcaacaaaagaggttacaacaaaattt
 cagctctgttaaagataatccggcagcgattaaagtacataaagatggctcgtttatttatc
 tgttatctaggtgattttaagacaactggaggcatatttgcgacaacagaaaaaggtgaa
 caaatagaagaaattatttctgatttaaatacagaatattgtattgatgacatggttttt
 gacagtaaaggcggattttatttctactgatttttagagggtattctacacaacctttgggc

5 ggtgtttactatgtagatccagactttaagacggttacgccaattattcaaaatatttct
 gtggcgaatggtattgctttaagtagcgatgaaaaagtgtatgggtaactgaaactaca
 actaatcgacttcaccgaatcgcattagaggatgatggcgtgactattgcaccatttga
 gcgacaataccatattattttacaggtcatgaaggaccggattcttgttattgatagt
 aatgataatttatatgtggctatgtatggccaaggacgtgtattagttttcaataagaga
 gggtatcctataggtcaaattttaatgccaggacgtgatgatggaaagatgttacgtaca
 acacatccacaatttatacctgggtacaaatcaacttataatttgtactaatgatattgaa
 aaccattctgaaggtggatctatgctttatacagttaatgggttttgctaaaggatatgag
 agttatcaatttcaataa

10

Sequence 1790

VFEGNIFKVNPAATKEVTTKFQSVKDNPAAIKVHKDGRLFICYLGDFKTTGGIFATTEKGE
 QIEEIIISDLNTEYCIDDVFDKSGGYFTDFRGYSTQPLGGVYVDPDFKTVTPIIQNIS
 VANGIALSTDEKVLVVTETTTNRLHRIALEDGVTIAPFGATIPYYFTGHEGPDSCCIDS
 15 NDNLVAMYQGQRLVFNKRGYPGQILMPGRDDGKMLRTHPQFIPGTNQLIICNDIE
 NHSEGGSMPLYTVNGFAKGYESYQFQ*

Sequence 1791

Contig_0652_pos_995_477,
 20 putative peptide of unknown function
 atgttagaaaacacatagattaaaagctagtgaagcctaatttgagttatacagatgaactt
 tatcaattgcatacaaaataaggttagctacaaagtatacacctaaaggatttcacagaat
 aaagtagcaaccaagattttattaaaggatggatgagggcattgggatgaatatcaattt
 ggttacttcattttaattatgagagataatcacgaagtagtggggatagcgggatttgag
 25 tatcgtacaattcatcaacaacagtttcttaatgcgtattatagaatctttccatcgtat
 actgggtgttggttagcttttgagtcaatggaggagattgcccgtcatttaaaaaagcat
 gataccataaacacaaaattaattcgaacaaatcaatataatacaaaattctattaaatta
 gcacaaaaactcggatataattatgatgctaactgggacgatgtaattaataaaggagat
 cgttggttttttaacctacaagcgttgataataactaa

30

Sequence 1792

MLETHRLKLVKPNLSYTDELYQLHTNKVATKYTPKGIHQNKVATQDFIKGWMRHWDEYQF
 GYFILIMRDNHEVVGIAGFEYRTIHQQQFLNAYYRIFPSYTGVLAFESMEEIARHLKKH
 DTITPKLIRTNQYNTNSIKLAQKLGYNVDANWDDVINKGDRCFFNLQALDNN*

35

Sequence 1793

Contig_0653_pos_4048_0,
 is similar to (with p-value 1.0e-21)
 >sp:sp|P37965|GLPQ_BACSU GLYCEROPHOSPHORYL DIESTER PHOSPHODI
 40 ESTERASE (EC 3.1.4.46) (GLYCEROPHOSPHODIESTER PHOSPHODIESTER
 ASE). >pir:pir|S37251|S37251 glycerophosphoryl diester phosp
 hodiesterase - Bacillus subtilis >gp:gp|Z26522|BSGLPTQ_2 B.s
 ubtilis glpT and glpQ genes for glycerol 3-phosphate permeas
 e and glycerophosphoryl diester phosphodiesterase. NID: g403
 45 371. >gp:gp|Z99105|BSUB0002_42 Bacillus subtilis complete ge
 nome (section 2 of 21): from 194651 to 415810. NID: g2632457
 . >gp:gp|AB006424|AB006424_43 Bacillus subtilis genomic DNA,
 70 kb region between 17 and 23 degree. NID: g3599592.
 gtgactttgatgaaatttcgacgtccaaatcaacatttccaaatcgtagcgcatagagga
 50 ttacctgaagattatcctgaaaatactattatcgcttatcgacatgcgctcatgttacat
 atagatatgttgaaattgatgtacattacacaaaagataaagaacttgctcgttatcat
 gatgatactatcgatcgtagctcaaatggtaaaggtaaggtttctgattttactttaaaa
 gaattaaaagcgtagatttttggtttctataaaggagagaaatttaaaggggagagtata
 ccgacttttgatgaagtgttagatttagcagataacttttcacaaaaattattaatagaa
 55 ataaaaagcctagtcagtagtccaaatattgaaaatattgattgttgataaattgaaggaa
 agacaaatatctaaatctaaagtgttttacaatcattcgattttgattgtgtgtaaaaaa
 ttgtcagcaatgaatttagattatgaattaggtttattaattagtaagaaaaaatattgg
 cacaagttaccaaaatttcaaaaaaattgccaaagttgctgattatgctaatacctaattat
 caaattg

Sequence 1794

VTLMKFRRPNQHFQIVAHRLPEDYPENTIIAYRHALMLHIDMLEIDVHYTKDKELVVIH
DDTIDRTSNGKGVSDFTLKLKALDFGFYKGEKFKGESIPTFDEVLDLADNFSQKLLIE
5 IKKPSQYPNIENMIVDKLKERQISKSVILQSFDFDCVKKLSAMNLDYELGLLISKKKYW
HKLPNFKKIAKVADYANPNYQIX

Sequence 1795

Contig_0653_pos_3981_3541,
10 putative peptide of unknown function
atgggttagacatgattttaaagttaaaactgaatggctaggtggacgcgaagaagtaggt
aaacttcgaggagatattatcaatgaaaatataatccatcccctcttcactcgggtggccaa
ggagaaggcacaatccagatgaattactagtaagtgcagcatcatcttggtacatcatt
tcactcgctgcaaacactagaaaaagagtggttttactaatgtaaaaattaatcaacagtca
15 ataggtacagcctcggttgaaaaataaaaaattttaaaatggaacgtattacacattatccc
tcaattaaagtaccctcttctcaaacagaaaagcttaaaagtattttagataaattatta
gtgattgcagataataattgtatgatatctaataacggaataatgtcattatttca
attgaacctaatttaataataa

20 Sequence 1796

MVRHDFKVKTEWLGGREEVGKLRGDIINENISIPSSLGGQGEGTNPDELLVSAASSCYII
SLAATLEKSGFTNVKINQQSIGTASFENKKFKMERITHYPSIKVPSSQTEKLKSILDKLL
VIADNNCMISNAIRNNVIISIEPNLI*

25 Sequence 1797

Contig_0653_pos_2142_667,
is similar to (with p-value 0.0e+00)
>gp:gp|Z99116|BSUB0013_19 Bacillus subtilis complete genome
(section 13 of 21): from 2395261 to 2613730. NID: g2634723.
30 >gp:gp|L47648|BACSERA_1 Bacillus subtilis phosphoglycerate d
ehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS, ypb
D, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (ypcA), y
pdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB, ypf
A, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA, yp
35 hA, yphB, yphC, NAD+ dependent glycerol-3-phosphate dehydrog
enase (glyc), yphE and yphF genes, complete cds. NID: g11461
95. >gp:gp|L47648|BACSERA_1 Bacillus subtilis phosphoglycera
te dehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS,
ypbD, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (ypcA
40), ypdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB,
ypfA, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA
, yphA, yphB, yphC, NAD+ dependent glycerol-3-phosphate dehy
drogenase (glyc), yphE and yphF genes, complete cds. NID: gl
146195.
45 atgatttcaacttatgatgctcttatcgtagcaagtcaaaccgaagtaacagagcgaatt
attaatgctgcaacaaatttgaaggctattgcaagagctggtgtaggtgtggataatatt
aatatagaagcagcgacttttaaagggtatttttagtaattaatgctcctgatggtaataca
atttctgctacagaacattcagtagctatgttgcttgcaatggcacgaaatattcctcaa
gcacaccaatctttacgtaacaaagaatggaatcgtaaagcatttagaggggttgaactt
50 tatggcaaaaccttaggtgttatcggtgctggtaggattgggttggcgctcgctaaacgt
gcgcagaggttccggtatgaaaatttttagcggttcgatccttatttaacagaagataaagcg
aagtcatttagatattcaaattgcaactggtgatgaaattgccgaaaaatccgactttgta
acagttcacacaccattaacacctaaaactcgaggaattgttggttcattctttctttaac
aaagctaaacaaaacttacaaatcataaatgttgccagagggggtattatagatgaaact
gcacttattgaagcattagataataacttaataagatcgtagcagctattgacgtatttgaa
55 catgaacctcctactgattccccctctcattcaacatgataaaattattgtcacaccacat
cttggcgctctactgtagaagcgcaagagaaggttgacgtctctgtatctgaagaaata
attgaaattctaactaaagggaatgttgagcatgctgtgaatgctccaaaaatggattta
agcaaagttgataaaacaactcaaagctttataggtttaagtacaactattgggtgagttt

gctattcagcttctcgatgggtgctccgagtgaaattaaagttaaataatgctgggtgactta
 gcgcaaaatgacactagtttaattacaagaacaattataacgaacatcttgaaagaagat
 ttaggtaatgaagtcaatattattaatgcattagcaataacttaaccaacaaggtgtcacg
 5 tataatatagaaaaacaaaagaacattctggcttttagtagttacattgagctagaacta
 gttaatgatcaagataaaatcaaaattggcgcaacggtattcgaggttttggccaaga
 atagtacgtattaatgattactcacttgattttaaacctaaccaatatcaattagtaaca
 tgtcataaagataaacctggtatagtaggacaaacaggcaacctattgggaagtcacgga
 attaatattgcgtcaatgactttaggacgtaacgatgctgggtggagatgctttaatgatt
 10 ctttctattgatcaacaagcaagtgaaggaagtataaaaaattttaaatgaaacaagcgga
 ttcaacaaaattattagcactaagtttaacaatttga

Sequence 1798

MISTYDALIVRSQTQVTERIINAATNLKVIARAGVGVDNINIEAATLKGILVINAPDGNT
 ISATEHSVAMLLAMARNIPQAHQSLRNKEWNRKAFRGVELYKTLGVIGAGRIGLVAKR
 15 AQSPGMKILAFDPYLTEDKAKSLDIQIATVDEIAEKSDFTVHTPLTPKTRGIVGSSFFN
 KAKQNLQIINVARGGIIDETALIEALDNNLIDRAAIDVFEHEPPTDSPLIQHDKIIVTPH
 LGASTVEAQEKVAVSVSEEIIEILTKGNVEHAVNAPKMDLSKVDKTTQSFIGLSTTIGEF
 AIQLLDGAPSEIKVKYAGDLAQNDTSLITRTIITNILKEDLGNEVNIINALAILNQQGV
 YNIEKQKKHSGFSSYIELELVNDQDKIKIGATVFAGFGPRIVRINDYSLDFKPNQYQLVT
 20 CHKDKPGIVGQTGNLLGSHGINIASMTLGRNDAGGDALMILSIDQQASEEVIKILNETSG
 FNKIISTKLTII*

Sequence 1799

Contig_0657_pos_695_1033,
 putative peptide of unknown function
 25 atgctaggatttgcaggggattgggatacagtcattataaagattcaaaatcgaacact
 gatgtagcttcaaaagagactcagacttccaataaaaaacactcatgaagatacaacttca
 caaggtaaaatgcaaaatcaagttaatagccaaacaaacgaagtatcaaattgggacatca
 actaaaacacttagtgaaaaagcaaagcagttaagagaagcttttaacgtcaatgatgag
 30 gaagctcaaatttttagcagatgaaatcgatagagcagatgtaataaagatggcacgatt
 acaacggatgaaatgacgcctacttattttcatatataa

Sequence 1800

MLGFAGGLGYSHYKDSKSNITDVASKETQTSNKNTHEDTTSQGMQNVNSQTNEVSNGTS
 35 TKTLSEKAKQLREAFNVNDEEAQILADEIDRADVNKOGTITTDDEMTPTFYFHI*

Sequence 1801

Contig_0657_pos_2041_2544,
 putative peptide of unknown function
 40 atggctacttaggacttcttttttaaatccctcaaattcccttcatttcttccctcattcat
 gtcattaaatcgattgatcacttcacctgtacttttgtttactacagcacaaaattttat
 aggagttccggctgcattctcttacaagggtgttcaacataatattcattagcattgcttct
 atctatatttgttttaaatcaagaacttggttgcttattaatccctttattaatgtaata
 gttcgagacaatttgttctgcttcttgagcggaatttgtttttgagctgtttttgttt
 45 ttgattactattagaatttgttacgctcccattagtttgcattatccgttttattttg
 ttcttcttttgattgctttctgtttttttttacatttcttgtccacatgctgttaata
 ttattgacgagagtgcataatgttcctaatagtttcaatttcattttatatccctccggtt
 aaaatgttgttaagttcacctaa

Sequence 1802

MVLRTSFFKFLKFLHFFLIHVIKSIDHFTCTFVYYSTKFYRSSGCISYKVFNIIFISIAS
 IYICFKFKNLLLINPFINVIVRDNLFCSLGNLFLSCLFLITIRICYAPISLHIIRFIL
 50 FFFWIAFCFFLHFLVHMLLILLTRVLMFLIVSISFYIPPFKMLLKFT*

Sequence 1803

Contig_0657_pos_5118_0,
 putative peptide of unknown function
 gtgcctatcaaaattaataatccagtggttagttaacacgagtttagagtgttaaggataat
 55 tttctaaaacttttggcggttccacaaatcaaccacgactaaatgtcccaaactcccaa

atgataagtagttggaatagtgataatgattaccggatcatttgaaaaatcgattaagttg
tttttaaaaaaggcggaatcctgcgttggttaaattgcggaaactgaagtgaataaaacttaa
aatagacctttacctatgccaaattttggaataaacgataaacatagacaaagtgtacca
aataattcagtgggcgatgctgtatatggctagatgtttaataagt

5

Sequence 1804

VPIKINNPVVVNTSLECKDNFLKLLAFHKSTTTKCPKPPKMISIGIVIMITGSFEKSIKL
FLKRANPALLNAETEVNKLKNRPLPMPNFGINDKHRQSVPNNSVAMLYMARCLIS

10

Sequence 1805

Contig_0657_pos_2384_1962,

putative peptide of unknown function

15

atgcaaactaatgggagcgtaacaaatttctaataagtaatacaaaaaacaaaaacagctcaa
aaacaaattttcgctcaagaagcagaacaaattgtctcgaactattacattaataaaggg
attaataagcaacaagttccttgaatttaaaacaaatatagatagaagcaatgctaatagaa
tattatgttgaacaccttgtaagagatgcagccggaactcctataaaattttgtgctgta
gtaaacaaaagtacaggtgaagtgatcaatcgatttaataatgacatgaatgaggaagaaatg
aaggaatttgaggaatttaaaaaagaagtcctaagtaccataattctggtgaagaaaaa
aataaaatgcaagaaaattcttcatcttccgaacaacaagaagtcacacaattctgttata
taa

20

Sequence 1806

MQTNGSVTNSNSNQKQKTAQKQ'ISAQAEQIVSNYYINKGINKQQVLEFKTNIDRSNANE
YYVEHLVRDAAGTPIKFCVAVNKSTGEVINRFNDMNEEEMKEFEFFKKRSPKYHNSGEEK
NKMQENSSSEQQEVHNSVI*

25

Sequence 1807

Contig_0658_pos_292_831,

is similar to (with p-value 2.0e-18)

30

>sp:sp|P37811|ATPD_BACSU ATP SYNTHASE DELTA CHAIN (EC 3.6.1.
34). >pir:pir|S39253|S39253 H+-transporting ATP synthase (EC
3.6.1.34) delta chain - Bacillus subtilis >gp:gp|Z28592|BSA
TPASE_5 B.subtilis (168) atpase genes for ATP synthase subun
its i, a, c, b, delta, alpha, gamma, beta, epsilon. NID: g4
33983. >gp:gp|Z99122|BSUB0019_181 Bacillus subtilis complete
genome (section 19 of 21): from 3597091 to 3809700. NID: g2
636029.

35

atggcaaaggtagcaaaaaaatatgccaaagcattatttgatgtcgctctagatacaaat
caactagatgttgtctatgaagatttagaacaattagccattcatcgtttgatttcac
aaacaacttaaagcaattgatagtaatccaagcttaactgcaaatcaacgtgaagaattt
gtagaaagagtttacaacgaagcaaatccatattgtgtaataactttaaaagtattagca
gataaccgacataatttcaattgtagagaattgttttaaatcattccaaaatttatataac
aaatactacaacaagattttgcaattattgaatcgacttacgagttaagcgaagetgaa
atatcaagaattgtagaacttatcaaaaagcaaaactgaattatcaaatgtaattgttaac
actaaaatcaatcaagatttaattggtggatttagagttaaggttggaactacagttatg
gatggttagtgtagaatgaccttggttcaattacaaagaaaatttgaaagagctaactaa

45

Sequence 1808

50

MAKVAKKYAKALFDVALDTNQLDVVYEDLETISHSSFDFIKQLKAIDSNPSLTANQREEF
VERVYNEANPYVVNTLKVLDNRHISIVENVFKSFQNLNKKYKQDFAIESTYELSEDE
ISRIVELIKKQTELSNVIVNTKINQDLIGGFRVKVGTVMGDSVRNDLVQLQKFERAN*

55

Sequence 1809

Contig_0658_pos_854_2365,

is similar to (with p-value 0.0e+00)

>sp:sp|P17674|ATPA_BACME ATP SYNTHASE ALPHA CHAIN (EC 3.6.1.
34). >pir:pir|F31482|F31482 H+-transporting ATP synthase (EC

3.6.1.34) alpha chain - *Bacillus megaterium* >gp:gp|M20255|B
ACATPA_7 B.megaterium ATP synthase i,a,c,b,delta,alpha,gamma
,beta and epsilon subunit genes, complete cds, and ORF. NID:
gl42553.

```

5  atggccataaaagctgaagaaatcagtgacattgcttcgctcacaattgaaaattatgag
   tcagaaatgtccggttacagatggttggtacagtactccaaattggtgacggtatcgcat
   attcacggacttaacgacgttatggctggtgagctagtagaattccataacggtgttctt
   ggtttagcacaaaaccttgaagaatctaattgtgggtgtggttattttaggaccatatgaa
10  gaaattagtgaaggtgacgaagttaaacgtactggccgaattatggaagtaccagtcgga
   gaggaatgataggaagagttgtaatcctcttggaacacccattgacggacaaggtcca
   atcaatgcgactaaaactcgtcctgtagagaaaaagcaactggcgtaattggatcgtaaa
   tctgtagatgaaccattacaacaggtatcaaagcaattgatgcttttagtaccaattggc
   cgtggtcaacgtgaattaatcattggtgaccgtcaaactggtaaaacaactgttgcaatt
   gattcaatcttaaaccaaaaagatcaagatacaatttgtatttatgttgcaataggtcaa
15  aaagattcaacagttcgtgcaaatgttgaaaaattaagacaagcaggtgctttagactac
   acaatcgtttgatctgcacccgagctgatccagcaccattactttatattgcaccttat
   tctggtgtaactatgggtgaagagttcatgtttaatggaaaacatgttcttatcgtttac
   gatgatttaactaaacaagcggcagcataccgtgagctatcattattattacgtagacca
   ccaggtcgtgaagcatatcctggggacgtgttctacttacacagtagattattagaaga
20  cctgcaaaacttaacgatgatcttgaggcggttcaattactgctttaccaatcattgaa
   actcaagctggcgatatctcagcatagcttccaacaaatgttatctcaattactgacgga
   caaatattcttacaatctgatttattcttctcaggtgttagaccagcgattaatgctggg
   caatcagtatctcgtgttggtgggttcagctcaaattaaagcgatgaaaaaagttgcagga
   acattacgttttagacttagcttcatatcgtgagttagaatcatttgcgcaatttggttct
25  gatttagatgaatttacagctaaaaaattagcgcgtggtgaacgtactgttgaagtatta
   aaacaagggtcaaaataaccactgcctgtagaacatcaagtacttattatttttgcttta
   actaaagggttacttagatgatattcctgtccaagatatcaatcgttttgagagggaattt
   aaccactgggctgagtcgaatgcaactgaattattaaatgaaattagagaaactggtgct
   ttaccagatgctgataaatttgattctgctatcacagaatttaaaaaaggatttaataaa
30  tcagaagaataa

```

Sequence 1810

```

MAIKAEELISALLRSQIENYESEMSVTDVGTVLQIGDGIALIHGLNDVMAGELVEFHNGVL
.GLAQNLEESNVGVVILGPYEEISEGDEVKRTGRIMEVPVGEEMIGRVVNPLGQPIDGQGP
35  INATKTRPVEKKATGVMDRKSVDPLQTGIKALDALVPIGRGQRELIIGDRQTGKTTVAI
   DSILNQKQDQDTICIYVAIGQKDSTVRANVEKLRQAGALDYTIVVSASAADPAPLLYIAPY
   SGVTMGEEFMFNGKHVLIYDDLTKQAAAYRELSLLLRPPGREAYPGDVFYLSRLLER
   AAKLNDDLGGGSITALPIIETQAGDISAYVPTNVISITDGQIFLQSDLFFSGVRPAINAG
   QSVSRVGGSAQIKAMKKVAGTLRLDLASYRELESFAQFGSDLDEFTAKKLARGERTVEVL
40  KQGQNNPLPVEHQVLIIFALTKGYLDDIPVQDINRFEEFNHWAESNATELLNEIRETGA
   LPDADKFDSAITEFKKGFNKSEE*

```

Sequence 1811

```

Contig_0658_pos_2498_3316,
45  is similar to (with p-value 3.0e-78)
   >sp:sp|P20602|ATPG_BACME ATP SYNTHASE GAMMA CHAIN (EC 3.6.1.
   34). >pir:pir|G31482|G31482 H+-transporting ATP synthase (EC
   3.6.1.34) gamma chain - Bacillus megaterium >gp:gp|M20255|B
50  ACATPA_8 B.megaterium ATP synthase i,a,c,b,delta,alpha,gamma
   ,beta and epsilon subunit genes, complete cds, and ORF. NID:
   gl42553.
   atgaagcaaataacaaaggcgatgaacatggtttctagttcaaaattacgtagagctgag
   aaaaatactaaatcatttagaccttatatggaaaagatgcaagatgctattacagctgta
   gctgggtcaaatagtagtacttctaatacatccaatgcttaaatctagagatattaaaagaagt
55  ggttacttagtaatcactagtgataaaggcttagccggtgcctatagtacaaatgtttta
   aaaagcttagtaaacgatatcaattctaaacccaacgacagtagtgaaatagctctaate
   gttttagggtcagcaaggtgtagatttcttcaacatagaggatgaaattgaaagttct
   ttagttgaagttccagatcaaccttcatttaaatctattcaatctatagctaaacatgct
   attgatttatttagcgaggaaaacatagatgaattgactatttattacagtcattatggt

```

5 agtgtcttagaaaataaacctgcaactaaacaagttttaccattatctcaagaagattca
 ggtcaaggacatggtcaaagtctctcatcacgaatttgaaccagataaagaatctatttta
 agcggtattttgccacaatacgttgaaagcttaatctacggtagacaatcttagatgcaaaa
 gctagtgaacatgcttcacgtatgacagcaatgagaaatgcttcagataatgcgacagaa
 ctaatcgatgattttatcattagaatacaatagagcgagacaagctgcgattactcaacaa
 attactgaaattggttggtgatcatcagctcttgagtaa

Sequence 1812

10 MKQITKAMNMVSSSKLRRRAEKNTKSFPRPYMEKMQDAITAVAGSNSTSNHPMLKSRDIKRS
 GYLVITSDKGLAGAYSTNVLKSIVNDINSKPNDSSSEYSLIVLGQQGVDFFKHRGYEIESS
 LVEVPDQPSFKSIQSIKHAIDLFEENIDELTIYSHYVSVLENKPKQVPLPSQEDS
 GQGHGQMSSYEFEPDKESILSVLPQYVESLIYGTILDAKASEHASRMTAMRNASDNATE
 LIDDLISLEYNRARQAAITQQITEIVGGSSALE*

15 Sequence 1813

Contig_0658_pos_3338_4750,
 is similar to (with p-value 0.0e+00)
 >sp:sp|P37809|ATPB_BACSU ATP SYNTHASE BETA CHAIN (EC 3.6.1.3
 4). >pir:pir|S39256|S39256 H+-transporting ATP synthase (EC
 20 3.6.1.34) beta chain - Bacillus subtilis >gp:gp|Z28592|BSATP
 ASE_8 B.subtilis (168) atpase genes for ATP synthase subunit
 s i, a, c, b, delta, alpha, gamma, beta, epsilon. NID: g433
 983. >gp:gp|Z99122|BSUB0019_178 Bacillus subtilis complete g
 enome (section 19 of 21): from 3597091 to 3809700. NID: g263
 25 6029.
 atgggaattggccgtgtaactcaagttatgggtccagtaatatagatgttcgttttgaaat
 aacgaagtacctgagataaataatgcattacacatcgaaagttcctaaagaagatggagcg
 cttcaattaacattagaagttgcacttcaactaggtgacgatgtagttcgtagaattgca
 atggactcaactgacggcgttcaaagaggaatggaagttaaagatacaggtagagacata
 30 agtgtacctgctggtgacgttaactctaggaagagtggttaacgtactaggagaaactatt
 gacttagatgaaaaaattgatgattcagtagcagctgacccatccatagacaagctcca
 ggattcgacgaattatcaacaaaagtagaaatcttagaaactggtattaaagtagtagac
 ttattagcaccttacataaaaaggtggtaaaattggattatttgggtggtgcccgtgtaggt
 aaaaccgtactaatccaagaacttatttaataacatcgctcaagaacacgggtggtatctca
 35 gtattcgctggtggttgggtgaacgtacacgtgaaggtaatgatctttactatgaaatgagt
 gacagtggtgttatcaagaaaactgcaatggtctttggtcaaatgaatgagccacctggt
 gcacgtatgctgttagcattatccggattaacaatggccgaatatttccgagatgaagaa
 ggccaagatgtgttattattcattgataacattttcagattcactcaagctggttcagaa
 gtttctgctgttattaggtcgtatgccatcagctgttggttatcaacctacacttgctaca
 40 gaaatgggtcaattacaagaacgtataagttcaacaaataaaggttcagttacatcaatt
 caagctgttttcgtaccagccgatgactatactgacctgcccagcaacaacgttcgca
 cacttagattcaacaacaaacttagagcgtaaattaacagaaatgggtatttatccagct
 gtagaccgcgttgcttctacatctagagcttttggaaccttcagtagtaggtcaagagcat
 tatgatgtggcagtgaaagttcaatctactttacaaaaatatagagagttacaagatatt
 45 attgcatctcttggtatggatgaattatcagatgaagataaaacaaactgtggaacgagca
 cgtagaattcaattcttcttatcacaaaacttccacgttgacagaacaatttactggacaa
 aaaggttcatatgtacctgttaaaacaacagttgcagacttcagagatatatttagatggt
 aagtatgaccatattcctgaagacgcattccgtttagtaggtagcatggaagacgtaatt
 gagaaagcaaaagatatgggtgttgaagtctaa

50 Sequence 1814

MGIGRVTQVMGPVIDVRFEHNEVPEINNALHIEVPKEDGALQLTLEVALQLGDDVVRTIA
 MDSTDGVQRGMEVKDTRDISVPVGDVTLGRVFNVLGETIDLDEKIDDSVRRDPIHRQAP
 GFDELSTKVEILETGKIKVVDLLAPYIKGGKIGLFGGAGVGKTVLIQELINNIAQEHGGIS
 VFAGVGERTREGNDLYEMSDSGVIKKTAMVFGQMNEPPGARMRVALSGLTMAEYFRDEE
 55 GQDVLLFIDNIFRFTQAGSEVSALLGRMPASVGYQPTLATMGQLQERISSTNKGSVTSI
 QAVFVPADDYTDPAATTFAHLDSTTNLERKLTMGIIYPAVDPLASTSRALEPSVVGQEH
 YDVAREVQSTLQKYRELQDIIAILGMDELSDEDKQTVERRRIQFFLSQNFHVAEQFTGQ
 KGSYVPVKTTVADFRDILDGKYDHIPEDAFRLVGSMDVIEKAKDMGVEV*

Sequence 1815

Contig_0658_pos_5233_5556,

putative peptide of unknown function

5 atgttaaaaaagtaatatccatttttataagtatctttattgcctcttttcgtgtacaatg
ttatacagtaaatcttacgaaggagcttaatatggattatcttggcaggttcgcaattggt
catttaaatcttacatggttgttatttgcgtagcttattgggctttaaattccataaaa
ttagaccaattttttaaaaagggtacccattacaagttcaagtttgatgattttatt
10 tctattttactgggtacggcagtcagtaactttatagttgattttattgcaatattcaact
caagtgaaataacttgataaaaataa

Sequence 1816

MLKSNIHFYKYLYCLFSCMTLYSNLTKEINMDYLGQFAIVHLILHVVCICVAYWALNSIK
LDQFFKKGYPLQVQVCMIFISILLGTAVSNFIVDLLQYSTQVKYLIK*

15

Sequence 1817

Contig_0658_pos_5666_6931,

is similar to (with p-value 0.0e+00)

>gp:gp|Z81356|BSATPC_4 B.subtilis atpC gene. NID: g1648848.
20 >gp:gp|Z99122|BSUB0019_173 Bacillus subtilis complete genome
(section 19 of 21): from 3597091 to 3809700. NID: g2636029.
atggataaaaatagtaataaatggtggaaatcgtttaacaggtgaagttaattgtgaagga
gctaaaaatgctgtattacctgtacttacagcgctcattacttgcttctgaaggacacagt
aaactagttaatgtcccagagtttaagcgatgttgaaacaatcaataatgtattatctaca
25 ctcaatgcaaatgtagagtatgataaagataaaaaatgcagttaaagttgatgcaactaaa
acttttaaatgaagaagcccttatgaatatgtgagtaaaatgcgagcaagcatcttggtt
atgggaccactacttctcgactagggcatgctattgttgctttgccaggaggatgtgctg
attggaacacgctctatagagcagcacattaaaggatttgaagctttaggtgcagacatt
catttagaaaatggaaatatttatgcaaatgctaaagatggattaaaggagcacatatt
30 catttagattttccgagtgtagggtgcaactcaaaatattattatggcagcgctccttggca
tcaggaaaatctatcattgagaatgtcgctaaagaacctgaaattgttgatttagctaat
tatattaatgaaatgggcggtaaaattacaggtgcaggtactgataccattacgatacat
gggtgtgaaaaactttacggtgttgaaatgcaattataccagatagaattgaagctgga
acgttacttatcgaggtgcaataactcggtgacatattttagctggtgcaattaaa
35 gaacatatggctagtttaatatataagcttgaagagatgggggtagatcttgaatattat
gaagaagggtataagagttacagcaaatggagatttaaatccagtagatgtaaaaacttta
ccacaccaggtttcccaactgatatgcaatctcaaatgatggctttattattaacagca
aatggacacaaaagtgtactgagactgtttttgaaaatagatttatgcacgttgacagaa
tttagaagaatgaatgcgaatataagtgttgaaaggaagaagtgtctaaaattgaaggaaaa
40 agccatttacaaggtgctcaagttaaagcaacagatttaagagctgctgcagccttaatc
ttagctggttttagttgcagagggaactacgcaagtgactgagttaaagcatctagataga
ggatacgtcaattttacatggaaaactaaaaagtctaggtgcaaacatagaacgtgtaaat
cgataa

45 Sequence 1818

MDKIVINGGNRLTGEVNVEGAKNAVLPLVTASLLASEGHSKLVNVPESDVETINNVLST
LNANVEYDKDKNAVKVDATKTLNEEAPYEYVSKMRASILVMGPLLARLGHAIVALPGGCA
IGTRPIEQHIKGFALGADIHLENGNIYANAKDGLKGAHIHLDFPSVGATQNIIMAASLA
SGKSIENVAKEPEIVDLANYINEMGGKITGAGTDTITIHGVEKLYGVEHAIIPDRIEAG
50 TLLIAGAITRGDIFVRGAIKEHMASLIYKLEEMGVDLEYEEGIRVTANGDLNPVDVKTL
PHPGFPTMQSQMMALLLTANGHKVITETVFENRFMHVAEFRMNANISVEGRSAKIEGK
SHLQGAQVKATDLRAAAALILAGLVAEGTTQVTELKHLDRGYVNLHGKLSLGANIERVN
R*

55 Sequence 1819

Contig_0659_pos_2951_4216,

is similar to (with p-value 0.0e+00)

>sp:sp|P00952|SYY_BACST TYROSYL-TRNA SYNTHETASE (EC 6.1.1.1)
(TYROSINE--TRNA LIGASE) (TYRRS). >pir:pir|A01179|SYBSYF tyr

osine--tRNA ligase (EC 6.1.1.1) - *Bacillus stearothermophilus*

5 atggcctaatgctttaatagaagattttaaagtggagagggttaatttatcaacaaacagat
gaagaggggtatcgagaattattaaataaggaacaggtcactttatattgcggcgagat
cctacagctgatagtttacatattggtcacttggtacctttttaacattaaagacgcttc
caagaacatgggcatcgctcctatcgtcttaattggtggaggtactggtatgataggagat
ccttctggaaagtctgaagagcgctgttacaacagaatcacaagttgaagctaacgtc
aaaggcctgtctaatcaaatgcatcgattatttgaatttggcagtgataaaggggcaaaa
ttagttaataataaagattggcttggctcaaatctcgttgattagtttcttagagattat
10 ggtaaacatgctggcggttaactatatgctaggaaaagattctattcaaacacgtttagaa
catgggtatctcttacagaatttacttatactattttacaagctattgattttggctat
ttaaatcgtgttagttaattgtaaaattcaagtaggcggatctgatcaatgggtaaatatt
acaagtgggtattgaattaatgcgacgaatgtatggacaaactgaggcatatggcctaaca
atcccattagtaactaaatcagatgggaagaaatttggtaaatcagagctcggagctgtg
15 tggtagatcctgaaaagacaagtcctatatgaattttatcaattttggattaatcaatct
gacgaagatgtaattaaattcttaaaatattttacttttttagaaaaagaagaattaat
cgattagaacaatcaaaaaatgaagcgctcacttacctgaagcacagaaagcactagcg
gaaaattgttacgaaatttattcatggtgaagcagctttaaaagacgctatacgtatttca
aaagcactgttttagtgagattttaaattcattatctgctaaagaacttaagaagggttt
20 aaagacgtacctcaagtaacgctatctacaaaaacgacaaatatagttgaagcacttatt
gaaacaggtattgcttcatctaaacgccaagcacgtgaagatgtaaacacgggtgcaata
tatattaatggtgaacgtcaacaatcagtcgattatgagttaagtagtaagaccttatt
gaagatgaaattacaataattcgacgaggaagaaaaaatatttcatggttaattaccaa
tcataa

25

Sequence 1820

MANALIEDLKWRGLIYQQTDEEGIEELLNKEQVTLYCGADPTADSLHIGHLLPFLTLRRF
QEHGHRPIVLIGGCTGMIGDPSGKSEERVLTQTESQVEANVKGLSNQMHRLEFEFGSDKGAK
LVNNKDWLGQISLISFLRDYKHHVGVNYMLGKDSIQTRLEHGISYTEFTYILQAIIDFGY
30 LNRELNCKIQVGGSDQWGNITSGIELMRRMYGQTEAYGLTIPLVTKSDGKKFKGKSESGAV
WLDPEKTSPEYFYQFWINQSDDEVIKFLKYFTFLEKEEINRLEQSKNEAPHLREAQKALA
ENVTKFIHGEAALKDAIRISKALFSGDLKSLSAKELKEGFKDVPQVTLSTKTTNIVEALI
ETGIASSKRQAREDVNNGAIYINGERQQSVDYELSSKDLIEDEITIIIRGKKKYFMVNYQ
S*

35

Sequence 1821

Contig_0659_pos_6502_5840,

putative peptide of unknown function

40 atgccttcagaattatggaatcgacacaaagttaaagatttttttgggaaaatcacaaact
tttttagatggcccttatgaaaataaggtagccagaaattttaccaaaggacagtttacg
gatgatactgcacaatcactactaatcattgatgcgctaaataaaaaatcattttgaacct
tcaaaaaaaatcatagcagacgaattgattgaatgggccaatgctacaaacgcattttaa
aataatattcttggcccaagtctaaagcagctttaaactgcaataatacaaggagaggat
tccaattatatacaaaaaatgcattaactaatgggtcggctatgagaattgctcctatt
45 ggtactttattttctaaagatcaaaaagtagagcttgtaaattatgttaaggaaattagt
gaagctactcacacaagtgatgttgcaatagctggagctagtatgattgcttatgcagta
acattggcagcgggaagataagaattggaaagaaattatagaaggcgttttggtattcac
gatatagctattaaagaaggagaaacaaactttttctgcatcaatagcagaaagattaaaa
ttagcagttcagtttgccaatcatttcgaaagtgaggaggaatatgttttcttttagtcat
50 tag

Sequence 1822

MPSELWNRNKKVDFFGKITTFLDGPYENKVARNFQKQFTDDTAQSLLIIDALNKNHFEP
SKKIIADELIEWANATNAFKNNILGPSSKAALTAIQGEDSQLYTKNALTNGSAMRIAPI
55 GTLFSKDQKVELVNVYKEISEATHTSOVAIAGASMIAYAVTLAEDKNWKEIEGVLDIH
DIAIKEGEQTFASIAERLKLAVQFANHFEESEEEYVFFSH*

Sequence 1823

Contig_0659_pos_5364_4411,

is similar to (with p-value 4.0e-60)
 >sp:sp|P39668|YYXA_BACSU HYPOTHETICAL PROTEASE IN ROCR-PURA
 INTERGENIC REGION (EC 3.4.21.-). >gp:gp|D78193|BACGNTZA_29 B
 acillus subtilis 36kb sequence between gntZ and trnY genes e
 5 ncoding 34 ORFs. NID: g1064780. >gp:gp|Z99124|BSUB0021_141 B
 acillus subtilis complete genome (section 21 of 21): from 39
 99281 to 4214814. NID: g2636442.
 gtgataaacatgcacaaatctacaaaccttgatgatttattcaacggtaaggcatctaaa
 tcaaaagaagcggaattggttccggtgtgatttatcaaataagtgaagggtccgcataat
 10 atcgttacaaataatcacgttggtgatggtgcttcggaaattaaagttcaactacataat
 tcaaaacaagtagatgccaaattaataggtaaagacgcctaacagatatgctgttcta
 aaaaataaagatacaaaaggaataaaagcaattcaatttgctaattcgtcaaaagttcaa
 acaggagatagtggttttgcgaatgggtaatcctctaggattagaatttgcaaatctgtt
 acatcaggaattatttcagctagcgaacgtacaattgacgccataacttctgctggtaat
 15 actaaagttaattgttttacagacagacgctgcaataaatcccggtaattcgggtggtgca
 ttagtgatatttaacggaaatctcgttggtatcaattccatgaaaattgcggcagcacia
 gtagaaggtatagggtttgctatacctagtaattgaagtttagagtaccatcgaacaactc
 gttaaacatggttaaaatcgaacgcccttcaatcgggtatagggtcttataaaatttaagtga
 attcctgaaaactatcgtaaaagaactacatactcataaagacaaaaggcgtttatgtagct
 20 aatgtagacagtgaaaatgccattaaaaagggtgatattattactggaatagatggtaaa
 caaataaaagatgatacagatttaagaacttatttatacagagagcaaaaaaccaggtgaa
 acggttactctaaaagttatcagagatggttaagacacaagacattaatgtaaaattaaaa
 aaacaagcatctgcattctgaatcatctcaatcacaagtcatttgctcaataa

25 Sequence 1824
 VINMQKSTNLDDLFGNGKASKSKEAGIGSGVVIYQISEGSAYIVTNNHVVDGASEIKVQLHN
 SKQVDAKLIGKDALTDIAVLKIKDKTGKIKAIQFANSSKVQTDGDSVFAMGNPLGLEFANSV
 TSGIISASERTIDANTSAGNTKVNVLQTDAAINPGNSGGALVDINGNLVGINSMKIAAAQ
 VEGIGFAIPSNEVRVTIEQLVKHKGKIERPSIGIGLINLSDIPENYRKELHCHKDKGVYVA
 30 KVDSENAIIKKGDIITGIDGKQIKDDTDLRTYLYESKKPGETVTLKVIDRGKTQDINVKLK
 KQASASESSQSQSQFAQ*

Sequence 1825
 Contig_0659_pos_2551_1646,
 35 is similar to (with p-value 3.0e-44)
 >gp:gp|AJ002293|SPAJ2293_1 Streptococcus pneumoniae pbplb ge
 ne, partial, beta-lactam resistant. NID: g2982645.
 atgcatttcaacaagtatcaaataactaactacagataaatataactaaatttgaacattta
 tataagaaggtcaaacatatatgtgtcgttaatttttttgggtggttttttgattgggttt
 40 attactactgtgtcattagtattatacttccaacaactaactaaagatgcttcgtcaata
 agtgatcgagaggttgaaagcaaaaatccttcataacctggcgatgagctaataaatcat
 aataatcaaattttagaagaatatgatcattcacaataacactcatagttggaccgaac
 catgtaaattcaaataattatacatgcacttacagcctctgaagatacattattttataaa
 cataacggtattatgcctaaagcacttttaagggcgatgcttcaagatatcacaattca
 45 aaccaatcttctggtggtagtactatcacgcaacaattagtaaaaaatcaagtgcctca
 aataaaaaaacttatagtcgtaaaagcaaatgaaattatcttggtacacgggtcgaaaat
 ttattatcaaaagatgaaatcatatatacgtattttaaataatcgctccatttggtcatgac
 tacaacggtgccaatataactggtatatcgtctgcttcataatagctctgttggtatacct
 gcaaaagatttgaatattgcacaatcagcttatctcattggcttactgcaaagtccatac
 50 ggctatacgccttatgacgaacacggcaaaagtaaaagccttaccatcttttaaaattgagc
 atgaaacgtaacaatacgtacttaaacgtatgcgtgttgaggaaaaatttctaacaacaa
 caatacgaacacgctaaaaaatacaatattaaacagcacttgctgaaacaatcgaaagac
 gaataa

55 Sequence 1826
 MHFNKYQILTDDKYTKFEHLYKKVKHICVVI FLVVFLIGFIILLSLVLYFQQLTKDASSI
 SDRELKAKILHIPGDELINHNQILEEYDHSQNTLIVGPNHVNSNIIHALTASEDTLFYK
 HNGIMPKALLRAMLQDITNSNQSSGGSTITQQLVKNQVLSNKKTYSRKANEIILATRVEN
 LLSKDEIIYTYLNIIVPFGHDYNGANITGISSASYSFLGIPAKDLNIAQSAYLIGLLQSPY

GYTPYDEHGKVKPYHLLKLSMKRQQYVLKMRVEGKISKQQYENAKKYNIKQHLLKQSKD
E*

Sequence 1827

- 5 Contig_0660_pos_7907_7290,
is similar to (with p-value 3.0e-52)
>sp:sp|P52996|PANB_BACSU_3-METHYL-2-OXOBUTANOATE HYDROXYMETH
YLTRANSFERASE (EC 2.1.2.11) (KETOPANTOATE HYDROXYMETHYLTRANS
FERASE). >gp:gp|L47709|BACYPIA_17 Bacillus subtilis (clone Y
10 AC15-6B) ypiABF genes, qcrABC genes, ypjABCDEFGHI genes, bir
A gene, panBCD genes, dinG gene, ypmB gene, aspB gene, asnS
gene, dnaD gene, nth gene and ypoC gene, complete cds's. NID
: g1146223. >gp:gp|Z99115|BSUB0012_183 Bacillus subtilis com
plete genome (section 12 of 21): from 2195541 to 2409220. NI
15 D: g2634478.
atgacagtgttaggatatgatagtactgttcaagttacattgaacgatatgattcatcat
ggtaaggctgttaaaagaggtgcttcagatacatttatagttgttgatagcctataggg
actgttggtttaagtgtatgaagaagatctaaaaaatgcacttaagctttatcaaaacacg
aatgctaacgctgtcaaaagtagaaggggctcatcttacatcatttattcaaaaagcaact
20 aaaatgggtatacctgttgtttctcacttaggtcttacacctcaaagtgtaggtgtaatg
gggtataaaacttcaaggggatacaaaagacagccgctatgcaacttatcaaagatgctaaa
gctatggaaactgctggtgcagtagtactgggttttagaagccatacctagtgtatttagct
cgagaaattagtgcgaactcactattccagttataggtataggggcaggaaaagataact
gatgggcaagtgttagtgatcatgatatgttaaattatgggtgttgatcgacacgctaag
25 tttgttaagcaatttgcagacttttcaagtgggtattgatggattaaggcaactgaaaaa
acaaataaatacctatga

Sequence 1828

- 30 MTVLGYDSTVQVTLNDMIHHGKAVKRGASDTFIVDMPIGTVGLSDEEDLKNALKLYQNT
NANAVKVEGAHLTSFIQKATKMGIPVVSGLGLTPQSVGMGYKLQGDTKTAAMQLIKDAK
AMETAGAVVLVLEAIPSDLAREISQQLTIPVIGIGAGKDTDGQVLVYHDMNLNYGVDRHAK
FVKQFADFSSGIDGLRQLKITNKYL*

Sequence 1829

- 35 Contig_0660_pos_7093_6287,
putative peptide of unknown function
atggatccgagtttgattttaccttatttatgggtacttttagtactttgtatttttagaa
ggattattagcagctgataatgcaattgtaatggcggaatggttaaacatctaccacct
aaacaacgtaaaaaagcacttttttatggcctattaggtgcattcatttttagatttatt
40 gctttatttttaataagtattattgcaaaccttctgggtgattcaagcagcagggtgctgtt
tacttaatctatatgtctattaaaaatttatggcaattttccatcagtc aaatgaaaaa
caccataaaagaaacaggagacgaacatcatttcgatgaaacaggcaacgaaaaagaagta
ggccctaaatctttttggggaacagtatttaaagttgaattcgctgatatcgcgtttgca
attgattcgatgcttgccgcattagccatagccgttacattaccaaaagttggcatacat
45 tttgggtggtatggacttagccaatttattgttatgttccttgggtggaatgatagggtgc
atcttgatgagatttgcagcaacttgggtttgtagaattggtgaataaatatccaggactt
gaagggtgctgcgtttgcaattgtaggttgggtaggtattaaacttgttataatggtaactt
gcacatcctgatattggcgttttaccagaagcatttccacatagtgtttatggcaaaaca
atcttctgggtagttatttagttggcttagttttaataggatgggttaacttcagcaattggc
50 aacaagaaaaaggtaatcaaaaataa

Sequence 1830

- MDPSLILPYLWLVLFLEGLLAADNAIVMAVMVKHLPKQKRALFYGLLGAFIFRFI
ALFLISIIANFWWIQAAGAVYLIYMSIKNLWQFFHQSNKHHKETGDEHHFDETGNEKEV
55 GPKSFWGTVFKEVFADIAFAIDSMALALAIATLPKVGIFHGGMDLGQFIVMFLGGMIGV
ILMRFAATWFVLLNKYPGLEGAFAIVGWVGKLVIMVLAHPDIGVLPFAFPHSALWQT
IFWVVLVGLVLIGWLTSAGNKKKGKGNQK*

Sequence 1831

Contig_0660_pos_5800_4319,
is similar to (with p-value 3.0e-49)
>gp:gp|AF000658|SPDNAARG_2 Streptococcus pneumoniae R801 tRN
A-Arg gene, partial sequence, and putative serine protease (
5 sphtra), SPSpoJ (spspoJ), initiator protein (spdnaa) and bet
a subunit of DNA polymerase III (spdnan) genes, complete cds
. NID: g2109442.
atgaataaatctaaagacgacgataataaaaattggagaagaaagtctacatgatgtgcga
gtttcaagtatacctccactttaccgcatcaaaataagtcaataaaagactatgatgat
10 tctggaacgaaagtaaacacataactaaattgacttcaaaggaaatctatgctaggcgta
aattctaatactatactgagcaagattcaagaagtacacaacatattcttcaaaacatagc
tattcacaacccaaaagataaaagacaacgataataactcaacaagcgcaatttcttaaaaaa
gaagacaaacaacgtaacagagccgaaaatataaaaaaagttaatgaatttaaacattg
gtggtagctttctttaagaacactggcctaaaatgttaattattatcggtattatagta
15 ttacttttaataatataatgccatattcactacagttaataaaaaatgatcatacaaatgat
agtgcatttaacggtacagctaaagatgaaacaacagcgatgaaaattgctgaaaactct
gttaagtcagttgtaactgtcgagaatgatttgtctaatagcacgactgtgtctgataac
aaaaatgaatctgataatgagataggtcaggtgtcgctcacaacaaaagtgggcgactct
atttatatttttactaatgcacacggtgtcggtgatcaagaaaaacaaaagtaacatat
20 ggtaatgataaatctgtgacaggttaaggtaattggtaagataaatggctctgatttagca
gttgtaaaagctaaagttgctgacgaaaatattaaaccaatgactatgggggatttctaat
aatattaaattggctgaacctatttttagttataggcaatcctctaggcacagactttaa
ggaagtgtttctcaaggtattgtgtctggactcaatcgatgtacctgtggacattgat
aaaaatgataattatgatgctttgatgaaagcttttcaaattgatgcgcctgtgaacca
25 ggaatttcaggtggtgctgtggtcgatagagacggttagactcataggtatagtttcttta
aaaattgatatgcataatgtagaaggaatggcttttgcgataacctattaatgatgtacgt
aagattgcgaagaatttagagcataaaggtaaaagtgaactacccaatactgaaatcaaa
atcaaaaatgttggtgaccttgacgattctgaacgtaatgcaatcaacttaccagctaaa
gtgaatcatgggtgtattaatcggtgaagtgaagaaaatggtttaggagacaaaagcaggt
30 ttaaaaaaaggtgatgtaatagtagaattagatggtaagaagattgaagataaacttacga
tatagacaagtcataatagtcattatgatgatcaaaaaacaattactgctaaaatttat
cgaaatgggtgcggagaaaaatattaaaatcaaattgaaataa

Sequence 1832
35 MNKSKDDDNKIGESLHDVRVSSDTSTLPHQNKSIKDYDDSGNESKQHTKLTSKESMLGV
NSNHTEQDSRSTQPYSSKHSYSQPKDKDNDNTQQAQFLKKEDKQRNRAENIKKVNEFKQL
VVAFFKEHWPKMILLIIGIIVLLILNAIFFTNNKNDHTNDSAFNGTAKDETTAMKIAENS
VKS VVTVENDSLNDTTVSDNKNESDNEIGSGVVYKKVGD SIYIFITNAHVVDQEKQKVTY
GNDKSVTGKVIKDKWSDLAVVKAKVADENIKPMTMGDSNNIKLAEPILVIGNPLGTD FK
40 GSVSQGIVSGLNRHVPVDIDKNDNYDALMKAFQIDAPVNPNGSGGAVVDRDGRLLIGIVSL
KIDMHNVEGMAFAIPINDVRKIAKELEHKGVNYPNTEIKIKNVGDLDDSERNAINLPK
VNHGVLIGEVKENG LGDKAGLKKGDVIVELDGKKIEDNLRQVIYSHYDDQKITAKIY
RNGAEKNIKIKLK*

Sequence 1833
Contig_0660_pos_4301_2943,
is similar to (with p-value 2.0e-42)
>sp:sp|P43440|NTPJ_ENTHR V-TYPE SODIUM ATP SYNTHASE SUBUNIT
J (EC 3.6.1.34) (NA(+)- TRANSLOCATING ATPASE SUBUNIT J). >gp
50 :gp|D17462|ENENTP_11 Enterococcus hirae ntp genes for Na⁺ -A
TPase subunits, complete cds. NID: g487271.
atgtcagttttaagtcagctgcttaaaaaatctagccctcaacaagggtataactctat
tatctattttgccatcgctcgttgctattttttattattaaatttgccatattgtcataaa
gggggttagggttaattccaattgatacactttttgtagctgtatcaggtattagtgttaca
55 ggattatcaccaattactatagtcgacacatactcaaccttttggtcaaattatcatactt
attattttgaacatcggtggtataggtgttatggctattggaacctgctttgggtgtt
ttaggcaaacatattggaattagagaacgacaactcattatgttagataataatagagat
actatgagtggaacggttaaactgatttttagaaattgtaagaacaatttttattattgag
tttataggtgcactcttactcgcttctattttttatcgggataatcctgatttaaaaaac

gcattaatgcaaggcatttttgtgtcagtatctgctacgactaacgggtggctcttgatatt
 actgggtgaatcttttagttccctatgcaaaagactattttgtacaaacaatagtaatgttc
 ttaattgtgttaggatctataggtttccagttttgttggagttaaagcatatataaag
 aatcgagtaactaatttttaggttctcattattcacaaaaataacgacaacaacgtattta
 5 tcttatttttattcggcggttattgttgtattaatattagaacatagcaatgcatttaaa
 ggattgagttggcatcaatcattatttatgcattgttccaatcatcaacgaccagaagt
 gcaggtttgcaaacgatagatgtgtcacatttcagcgacgcaacaatatgttaatggga
 ttattgatgtttataggatcatctccagttctgtaggaggtggaatcagaacaacaaca
 tttgctatttttaattttgtttgttattaattttaataataactggtgacaaaacaggatt
 10 aaaattttcaacagagaagtacatattatggatgtacaaagatcatttgccgtatttact
 atggcgctcatttaattacatttattagcatgattattatctgccactgaacaaggcaag
 ttgtcctttttacaaattttctttgaagtaatgtctgcgttttggtacatgtggtttaagc
 ttaggtgtgacaagcgatgtcaacgacattaccaaggctgtattaatgatattaatgttt
 ataggtcgtgtaggacttatttcattcattattatgttagctggacgtaaagaaccgaa
 15 aatatagctatcctaagaacgtatacaaatggatag

Sequence 1834

MSVLSQLLKKSSPQQGIILYYLFAIVVAFLLNLNLPYVHKQGVENPIDTLFVAVSGISVT
 GLSPITIVDTYSTFGQIIILILNIGGIGVMAIGTMLWVVLGKHIGIRERQLIMLDNNRD
 20 TMSGTVKLILEIVRTIFIIEFIGALLLAFYFYRDNPDNLKALMQGIFVSVSATNNGGLDI
 TGESLVPYAKDYFVQTIWMFLIVLGSIGFPVLELKYIKNRVTNFRFSLFTKITTTTYL
 FLFLFGVIVVLILEHSNAFKLSWHQSLFYALFQSSTTRSAGLQOTIDVSHFSDATNIVMG
 LLMFIGSSPSSVGGGIRTTTFAILILFVINFNNTGDKTGIKIFNREVMHMDVQRSFAVFT
 MASLITFISMIISATEQGLSFLQIFFEVMSAFGTCGLSLGVTSDVNDITKAVLMILMF
 25 IGRVGLISFIIMLAGRKEPEKYSYPKERIQIG*

Sequence 1835

Contig_0660_pos_2849_1509,
 putative peptide of unknown function

atgaatggtacatgttacactaaattcaaggagaggagtggctctatggaaaaaatgaa
 aataaaatagttgatgtgattgcaacttctgatatgcatagtcacttcttaaatggatg
 30 tatggctcaaacattttatcggtctggtacttatgttaatgaagctcgaaaaaatatgaa
 aatgtcatattactagatagtgggggaagtttggcggtatcacttgacgattttattac
 gccgtttagcaccatataaacgtcatcctatgattaagttgatgaatgcaatgcagtat
 35 gatgctagtgggaattagccctaataaattttggattatcttttttgacacgttca
 gttgcactctcaaggtttccttggtatcagcaaatatagagtatactgttactagagag
 ccatatttttctacgccttatacaatcaaaatgtattcagatttgaaaattgctatcgta
 gttttaacatcagatggattaatgaagaacgagtacgcagaaatgaagaagatgtctgt
 attgagaagactttggtttcagctaaacgttggttagatatacatgaagttgaagaa
 40 ccagacttcttattgttatttatcatggtgggttaaataaaattagtagtgccaataaa
 agaaatgaaaaaatgcaaacgaagctgaaaaaattatggaagaacttggtgttattgat
 gtaattattaccgctcatcaacatcaaacagtagttggaaaagatcatggaactatatat
 gttcaagcaggtcaaaatgctgaggaattagtacatctttcaattaaatttaagaaacgt
 acaacttcttatgagattgagcacatcgactcaaaagttattgacttaaattgattaccat
 45 gaagatgagcaattattaaaagaaacatattatgatcgtaaggcagtcacaaactgggca
 aattcagtagtttcaaacaaaaacaatggcttaacagttcaatgtattgaagatattatt
 tgtaagccgcacatccttttactcaattattacatgatgcaattagattagcctataattat
 gatatttcttgtgtgcatatacctaagaatggtaggaagggttaaaaggaaactataaga
 aatagagatatatacgatgcataccctcatccagataaacctatagatatcactgtcaaa
 50 ggtaaaaatatcaaagatatacttgaatacagttatgcgcatttgatttttaataagcga
 caatttattagaacggaggttaaattacaattttttgtttgattaaatttaactattat
 tcatttggacatctatttttaa

Sequence 1836

MNGTCYTKFKERSGSMENKIVDVIATSDMHSHFLNGDYGSNIYRAGTYVNEARKNNE
 55 NVILLDSGGSLAGSLAAFYAVVAPYKRHPMIKLMNAMQYDASGISPNEFKGLSFLTRS
 VALSRFPWLSANIEYTVTREPYFSTPYTIKMYSDLKIAIVGLTSDGLMKNEYAEMEEDVC
 IEKTLVS AKRWIRYIHEVEEPDFLIVYHGGLNKKISSANKRNEKNANEAEKIMEELGVID
 VIITAHQHQT VVGKDHGTIYVQAGQNAEELVHLSIKFKKRTTSYEIEHIDSKVIDLNDYH

EDEQLLKETYDRKAVKHWANSVVSNNGLTVQCIEDIICKPHPTQLLHDAIRLAYNY
DISCVHIPKNGEGLKGTIRNRDIYDAYPHDPKPIDITVKGKNIKDILEYSYAHIDENKR
QFIRTEVKLQFICLIKFNYYSFGLHF*

5 Sequence 1837

Contig_0660_pos_1463_1002,

is similar to (with p-value 2.0e-48)

>gp:gp|Y17554|BLY17554_2 Bacillus licheniformis arcA, arcB,
arcC and arcD genes. NID: g3687415.

10 atgaatttccatcttgtttgtcctaaagaactcaatccgacagaagaattattaaatcgt
tgcaacgtattgagcgggaaatggcggttaacattttaataacagatgatattgataaa
ggcgtgaaagattctgatgttatttatacagatgtttgggtatcaatggcggaacctgat
gaagtatggcaagaacgccttaaaacttttaaaaccatatcaagttaaccaagcattatta
gaaaaaacaggcaatccaaatgttatttttgaacattgtttaccttctttccacaatgca
15 gaaactaaaattggtcaacaaatttatgaaaaatatggcattagtgaatggaagtcaat
gatgatgtcttcgaaagcaaagcttctgtagtattccaagaagctgagaatagaatgcat
acaattaaagcgggtcatggttagcaacttttaggagaattctaa

Sequence 1838

20 MNFHLVCPKELNPTEELLNRCERIATENGGNITDDIDKGVKDSVDIYTDVWVSMGEPD
EVWQERLKLKPYQVNQALLEKTGNPNVIFEHCLPSFHNAETKIGQQIYEKYGISEMEVT
DDVFESKASVVFQEAENRMHTIKAVMVATLGEF*

Sequence 1839

25 Contig_0661_pos_1122_2249,

is similar to (with p-value 0.0e+00)

>gp:gp|AF009352|AF009352_2 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
30 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_69 Bacillus subtili
s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: g2635827.

atgaccatagacaltgaatcaggagactttattgcatttattgggacaagtggtagcgggt
35 aaaacaactgcccttagaatgattaatcgtatgattgaatctacagagggagaaattacc
attgacggtaaaaaatatcaaagagcttaatccggttgagcttcgctgcagtatcggttat
gtcatacaacaaatcggttaatgccacatatgacagtgaagagaatattgttctcgta
ccaaagttattaaaatggtcacaagagaaaaaggatgagaaagcgaaagaacttatacgc
ttagtagatttaccagaagaatatttagatcgatatccttcagaattatcaggtggtcaa
40 caacaacgtattggtgttgtaagagcactgcagctgaacaagatattattttaatggat
gaaccggttgggtgactcgatccaatcacaagagatacattacaagacttagtcaaaaaa
ttacaacaacaattaggaaagacattcattttgttacacatgatatggatgaagcaatc
aaacttgagataaaaatatgtattatgacaaatggacaggtgattcaatatgacacgcca
gataatattttacgtagtccagcgaatgatttcgttagagactttattggtcagaatcgc
45 ttaattcaagatagacctaatatccgtacagttaaagatgcatgattaaaccggtgaca
gtacatggtgaccgttctcttaatgatgacggtgaatattatgagagagaaaaagatcgat
acgatatttgttgcggcaatgatgagcatttattgggttatttagatattgaagatatt
aacgaaggattaaagacatcataaagaacttatagatacgatgcaacgagatatttataga
gtacgtattgatagtaagttacaagattctgttcgtacaattcttaaacgtaattgtacgt
50 aatgtaccggttgttgacagtgataataaaacattattaggccttgtcaccgagctaac
cttgttagacattgtttatgacagttattggggagagttagaatcgggtaacaatgataat
cattctgggattgttgaacccgaggtccacaggagttgagacaccatga

Sequence 1840

55 MTIDIESGDFIAFIGTSGSGKTTALRMINRMIESTEGETIDGKNIKELNPVELRRSIGY
VIQQIGLPHMTVKENIVLPKLLKWSQEKKDEKAKELIRLVDPPEYLD RYPSELSSGGQ
QQRIGVVRALAAEQDIILMDEPFGALDPITRDTLQDLVKKLQQQLGKTFIFVTHDMDEAI
KLADKICIMTNGQVIQYDTPDNILRSPANDEFVRDFIQNRLIQDRPNIRTVKDAMIKPVT
VHVDRSLNDAVNIMREKRVDTIFVVGNDHLLGYLDIEDINEGLRHHKELIDTMQRIYR

VRIDSKLQDSVRTILKRNVRNPVVDSNKTLLGLVTRANLVDIVYDSIWGELESGNNDN
HSGIVEPESTGVETP*

Sequence 1841

5 Contig_0661_pos_2336_2881,
is similar to (with p-value 2.0e-44)
>gp:gp|AF009352|AF009352_3 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
10 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388. >gp:gp|Z99121|BSUB0018_68 Bacillus subtili
s complete genome (section 18 of 21): from 3399551 to 360906
0. NID: g2635827.
atgattgtagcgggttccttttaggcattttattatctaaaaaggaaaagctttcgaaagt
15 tcattgacgatagctggcgttttacaacgatacctacattagcggatttagctttaatg
attccattatttgggtgtaggaaaaacacctgcaattatagcgttatttttatatgtatta
ttaccaatttttaataatacagattataggcattcaaaatatagattccaaccttagagaa
gcaggacgtagtagtggaatgactaactttcaattgatgaaagatgtggagttgccactc
gcattaccattaataacttagtggaattaggtgtcttctgtctacgttattagttgggca
20 acattggcaagttatgttgggtgctggtggttttaggtgattttatctttaacggattagcg
ctatttgaaccgagtagtattactgcaactattctgtcactgctattgcacttatt
gtagattatgtttatcactgattgaaaaatgggtgtacctaaggattaaaagtttcc
agataa

Sequence 1842

25 MIVAVPLGILLSKEKLSKVSLTIAGVLQTIPTLAVLALMIPLFGVGKTPAIIALFLYVL
LPILNNTIIGIQNIDSNLREAGRSMGMTNFQLMKDVELPLALPLILSGIRLSSVYVISWA
TLASYVGAGGLGDFIFNGLALFEPMSMIITATILVTAIALIVDYVLSLIEKWVVPKGLKVS
R*

Sequence 1843

30 Contig_0661_pos_2900_3859,
is similar to (with p-value 3.0e-93)
>gp:gp|AF009352|AF009352_4 Bacillus subtilis osmoprotectant
transport system OpuC including ATPase (opuCA), transmembran
e protein (opuCB), osmoprotectant binding protein precursor
35 (opuCC) and transmembrane protein (opuCD) genes, complete cd
s. NID: g2271388.
atgaagaattataaaaaatatctaattgttttagtattatgtctaacagtgttatctgga
40 tgtaacttaccgggttttaaaaaatagtcattcagatgacgatgtagaatcacaagttta
ggaacaagtgaatcacaattatatcgacatgctgagattacttattgaacatgataca
aaaggtgaaattaaacctacctaattaataatttaggttcaagtacgattcaacataat
gctgtcacaagtggccaagctaataatgtcaggtacgcgttatacaggcactgacttaaca
ggggcgttgaaagaagatccgattaaagatcctaaaaaggccatgaaagcgactcaagaa
45 gggtttaaaaagaaatacaatcaaacccttcttcaattcatatggtttgaaaacacatat
gcattgatagtaacaaaagagacagctaagaaatatcatttagaaaactgtttcagactta
gaaaagcatgctaaagatttaagagtaggtatggatagttcatggatggaccgtaaaggt
gacggttatccagcatttaaaaaagaatatggctatagtttcggtacggttaagacctatg
caaattgggtcttctgtctatgatgcactgagttctggcaaattagacgttgccgtaggttat
50 tctacagatggacgtattttcagcatatgatttaaaagtgttgaaagatgatcgatcgtatc
ttcccaccatacgtatgccagtcacttgcacatcagatcaattgttaaaagagaagccagag
ctcaaaccgattgttaaaaaattagaaggtaagatatcaacagaaacaatgcaagaatta
aattatcaagctgatggttaaaagggaagagcctgcaacagttgcagaggacttcttgaaa
aaacatcattattttgaagacgatgacaataaaaaagataaacagaaaggtggtcaataa
55

Sequence 1844

MKNYKYLIVLVLCLTVLSCNLPGLKNSHSDDDVRITSLGTSESQIISHMLRLLEHDT
KGEIKPTLINLGSSTIQHNAVTSQGAMSGTRYTGTDLTGALKEDPIKDPKKAMKATQE

GFKKKYNQTFNNSYGFENTYALIVTKETAKKYHLETVSDLEKHAKDLRVGMDSSWMDRKG
 DGYPAFKKEYGYSFGTVRPMQIGLVYDALSSGKLDVAVGYSTDGRISAYDLKVLKDDRRF
 FPPYDASPLASDQLLEKPELKPIVKKLEGKISTEQMQELNYQADGKGKEPATVAEDFLK
 KHHYFEDDDNKKDKQKGGQ*

5

Sequence 1845

Contig_0661_pos_3862_4509,

is similar to (with p-value 3.0e-47)

>gp:gp|AF009352|AF009352_5 Bacillus subtilis osmoprotectant
 10 transport system OpuC including ATPase (opuCA), transmembran
 e protein (opuCB), osmoprotectant binding protein precursor
 (opuCC) and transmembrane protein (opuCD) genes, complete cd
 s. NID: g2271388. >gp:gp|299121|BSUB0018_66 Bacillus subtili
 s complete genome (section 18 of 21): from 3399551 to 360906
 15 0. NID: g2635827.

atggaagggaatttaatacaacaactcgttcactactatcaaataaactttggctaccta
 tgggaattgtttgttaatacattgttaattgtctgtttacggtgttttactagcatgttta
 gtgggcataacctcttggatcatcattgcacgatttggtaaattatcaggtgttattatc
 actattgcgaatattattcaaaactgttccgggtattgctatgttagcgatattaatgctt
 20 agtatgggacttggatgaatacagtcatttttactgtgttcttatatgccttgcttcct
 attattaaaaatacgtatacaggaattaatgaagttgacccaaatattaaagatgctgga
 aaagggatgggtatgacgcgtaaccaagtattaaactatgattgagttacctctgtcactt
 tcagttattataggtgggtattcggattgcccttgtttagctattgggtgctgtagctgta
 ggttcattttattgttgcaccaacattaggtgatattgtgattagaggtacaaatgcaact
 25 gatggaacactattttattctagcaggtgccatacctatcgttatcattgtcatacttata
 gatgtttttattacgtttattagagaaaaagctagatccagctacgtaa

Sequence 1846

MEGNLIQQLVHHYQMNFGYLWELFVNHLMSVYGVLLACLVGIIPLGIIARFGKLSGVII
 30 TIANIIQTVPMIAMLAILMLSMGLGMNTVIFTVFLYALLPIIKNTYTGINEVDPNIKDAG
 KGMGMTRNQVLTMIELPLSLSVIIGGIRIALVVAIGVVAVGSFIGAPTGLDIVIRGTNAT
 DGTLFILAGAIPIVIVILIDVLLRLLEKKLDPAT*

Sequence 1847

35 Contig_0661_pos_5935_5594,

is similar to (with p-value 8.0e-28)

>sp:sp|P44023|YFCC_HAEIN HYPOTHETICAL PROTEIN HI0594. >pir:p
 ir|E64010|E64010 hypothetical protein HI0594 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32741|U32741_2 Haemophilu
 40 s influenzae Rd section 56 of 163 of the complete genome. NI
 D: g1573582.

gtgcaacatatgagtgggcctttatttatcattgttctgtcttttatctttttctgttta
 ggatttatcgtgccgtcctcatcaggattagcagtagctatctatgcctatctttgcgcca
 ttagctgatacagtaggtataccaagatttgttattgttacaacatatcaattcggtcag
 45 tatgcaatgttgttcttagcgcctactggacttgaatggcaacacttcaaagtgttaaac
 atgcgctactcacactggttacgtttcgtatggcctgttgcgcgtttgttttaataattt
 ggtggaggcttacttattacacaagttttaataatactcataa

Sequence 1848

50 VQHMSGPLFIIVLLFIFFCLGFIVPSSGLAVLSMPIFAPLADTVGIPRFVIVTTYQFGQ
 YAMFLAPTGLVMATLQMLNMRYSHWLRVFWPVVAFVLIFGGGLLIQVLIYS*

Sequence 1849

Contig_0661_pos_4460_4065,

55 is similar to (with p-value 2.0e-30)

>gp:gp|AF009352|AF009352_5 Bacillus subtilis osmoprotectant
 transport system OpuC including ATPase (opuCA), transmembran
 e protein (opuCB), osmoprotectant binding protein precursor
 (opuCC) and transmembrane protein (opuCD) genes, complete cd

s. NID: g2271388. >gp:gp|299121|BSUB0018_66 *Bacillus subtilis* complete genome (section 18 of 21): from 3399551 to 3609060. NID: g2635827.

5 atgacaatgataacgataggtatggcacctgctagaataaatagtgttccatcagttgca
tttgtaacctctaatacacaatatcacctaattgttggtgcaccaataaatgaacctacagct
acgacaccaatagctacaacaagggaatccgaataccacctataataactgaaagtgc
agaggttaactcaatcatagtttaatacttgggttacgcgctcataccatcccttttccagca
tctttaatatgttggtcaacttcatttaattcctgtatacgtatttttaataataggaagc
aaggcatataagaacacagtaaaatgactgtattcataccaagtcccataactaagcatt
10 aatatcgctaacatagcaataaccggaacagtttga

Sequence 1850

MTMITIGMAPARINSVPSVAFVPLITISPNGAPINEPTATTPIATTRAIRIPPIITESD
RGNSIIVNTWLRVIPFPASLIFGSTSLIPVYVFLIIGSKAYKNTVKMTVFIPSPILSI
15 NIANIAITGTV*

Sequence 1851

Contig_0661_pos_2553_2143,
putative peptide of unknown function
20 atgcctataatcgattattttaaattggtaataatacatataaaaaataacgctataatt
gcagggtgttttctacaccaaataatggaatcattaaagctaataaccgctaattgtaggt
atcggttgtaaaacgccagctatcgtaacgactttcgaaagcttttcttttttagat
aataaaatgcctaaaggaaccgctacaatcattgcaataaccaatgccactatcgatatg
tacaatgttcgaatgtctttgcaataacataccactatgttccgaaataaactttatc
25 atgggtgtctcaactcctgtggactcgggttcaacaatcccagaatgattatcattgttac
ccgattctaactctcccaaatactgtcataaacaatgtctacaagggttag

Sequence 1852

MPIIVLFKIGNNTYKNNAILIAGVFPTPNNGI IKANTANVGIVCKTPAIVNDTFESFSFLD
30 NKMPKGTATIIAITNATIDMYKCSNVFANNIPLCSEINFIMVSQLLWTRVQQSQNDYHCY
PILTLPKYCHKQCLQG*

Sequence 1853

Contig_0664_pos_5849_5160,
35 is similar to (with p-value 1.0e-55)
>sp:sp|P52998|PANC_BACSU PANTOATE--BETA-ALANINE LIGASE (EC 6
.3.2.1) (PANTOTHENATE SYNTHETASE) (PANTOATE ACTIVATING ENZYME).
>gp:gp|L47709|BACYPIA_18 *Bacillus subtilis* (clone YAC15-6B) ypiABF
genes, qcrABC genes, ypjABCDEFGH genes, birA gene, panBCD genes,
40 dinG gene, ypmB gene, aspB gene, asnS gene, dnaD gene, nth gene
and ypoC gene, complete cds's. NID: g1146223. >gp:gp|299115|BSUB0012_182
Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220. NID: g2634478.
45 gtgaatcccttgacgttttgacctaattgaggattttgatgcttatccacgtcaactcgat
gatgatgtggctgcagtaaaaaagttacaagtggattatgttttccatccgagtgtagat
gaaatgtatccagaagaattaggtattcatctgaaagttggacacttggcacaagtatta
gagggagcacaagaacctggacacttcgaaggtgttggtgaccgtggtcaaaaactattt
aatattgtgcaaccagattatgcctattttgggaaaaaggatgcacaacaattagctatt
50 gttgaaaaagatgggttaaagactttaattcttctgtacatgttatcggtattgatatcgta
agagaaaaagatgggttagccaaaagctctagaaatatttacttgacctctgaagaacga
aaagaggcaaaacatttatatacaagctctacgcttagcaaaagaatttgatgaagcgggt
gaacgagatagcaatgagattataggtcaaatcgctgcgtatttaacaaaaaatattagt
ggacatattgatgatttgggtatttatagttatccaaatcttatacaacaatcaaagatt
55 catggacgaatattcatatcattggcagttaaattttctaaagcaagattgatagataat
ataattattggagatgactatattgattag

Sequence 1854

VNPLQFGPNEDFDAYPRQLDDVA AVKKLQVDYVFHPSVDEMYPEELGIHLKVGH LAQVL

EGAQRPGHFEGVVTVVNKLFNIVQPDYAYFGKKDAQQLAIVEKMKVDENLPVHVIGIDIV
REKDGLAKSSRNIIYLTSEERKEAKHLYQSLRLAKNLYEAGERDSNEIIGQIAAYLNKNIS
GHIDDLGIYSYPNLIQQSKIHGRIFFISLAVKFESKARLIDNIIIGDDYID*

5 Sequence 1855

Contig_0664_pos_5152_4781,
is similar to (with p-value 4.0e-35)
>sp:sp|P52999|PAND_BACSU ASPARTATE 1-DECARBOXYLASE PRECURSOR
(EC 4.1.1.11) (ASPARTATE ALPHA- DECARBOXYLASE). >gp:gp|L477
09|BACYP1A_19 Bacillus subtilis (clone YAC15-6B) ypiABF gene
s, qcrABC genes, ypjABCDEFGH genes, birA gene, panBCD genes
, dinG gene, ypmB gene, aspB gene, asnS gene, dnaD gene, nth
gene and ypoC gene, complete cds's. NID: g1146223. >gp:gp|Z
99115|BSUB0012_181 Bacillus subtilis complete genome (sectio
n 12 of 21): from 2195541 to 2409220. NID: g2634478.
atgaactcaaaaatccatagagctagagttacagaatctaatttaaattacgttggaagc
ataacaatagatgccaatatattagatgcggttgatattttacccaatgaaaagggtgct
attgttaataataataatggtgctcgatttgaacatatgtcattgctggagaacgtggt
agtggaaagatgtgttttaaatggcgcggcttcaagactagttgaagttggagacgtcatt
20 attattatgacatatgcacaattaaatgaagatgaaatggtagatcactcaccaaaagta
gctgtgttaaatgaaaataatgaaattatagaaatgataaatgagaaagaaatacgata
tcaaatgtataa

Sequence 1856

25 MNSKIHRARVTESNLNYVGSITIDANILDAVDILPNEKVAIVNNNGARFETYVIAGERG
SGKMCNLNGAASRLVEVGDVIIIMTYAQLNEDEMVDHSPKVAVLNENNEIEMINEKENTI
SNV*

Sequence 1857

30 Contig_0664_pos_4363_3353,
putative peptide of unknown function
atgcattgtccgaattgcgtaattccaatagaagatgatgatttattctgtggtgaatgt
ggacataaaaataaagtcgacatccacagtcagtcagaaatgcagaaagtgaataacaaaa
gctgaaaaagaatgatgaagaacaaaacataacatcaaataataaagaaaaaacgcagcg
35 actcatcaaaatggtgattcaacatctcatgatgagactagatcaaataaaaatgacgta
gcagattcaacattacaatctaagcagtcacataatgatcccaacaatcaaattctctct
acataccatcaaagacctcaacatcgagaaattcctcaaaatcaacacaatcacgatcaa
cagcaaagtcaaataaggtcaacaagctaacaagtaacaaatgaaagtaaagggtttcttt
aaaagtgcatttactgcacctgataaaatcattcaaaactaatcatgttttcagttttaa
40 ttattattatcattattattaataatcggttttattgtattagcaattttactcgcttccgta
ataccagttgagattggtattttcggtactacaagaggaagtttggttaacgagtatcatt
tttggtattattctatttttggtgtcatagtaggtgcaatatttggtggttacacgttta
gtagtttagacaacctattgcatttaaaaaagattatcagactatgtgtaattaatagt
gtttcgtttagcaattttaattatttctgtaattttaatttagcagaatcatacagcttt
45 ggcggaagtatagcattattgtctttattattatttattgcttctggtatttatctaatt
gcgaagtatagcactggtaatcaacaagaatatccagcttttatggtgtgattatttat
atcattattttgttcttatttatttcgtatttttggggaggcattttccatcaaatttt
ggtgattttatagaagaactaggggattttattgaaggagggaacttattaa

50 Sequence 1858

MHCPNCGNPIEDDDLFCEGCGHKISRHPQSVRNAESEITKAENDEEQNITSNNKENNAA
THQNV DSTSHDETRSNENDVADSTLQSKQSHNDTQQSNLSTYHQRPHREIPQNQHNDQ
QQSQIGQQAKQVTNESKGFFKSAFTAPDKIIQTNHVFSFKLLSLLIIGFIVLAILLASV
IPVEIGIFGTTRGSLVTSIIFGIILFLVVIVGAIFGLTRLVVRQPIAFKKVLSYVLINS
55 VSLAILIISVILILAESYFSGGSIALLSLLFIASGIYLIKYSTGNQTRISSFYGVIIY
IIILFLFIRIFGEAFFHQIFGDFIEELGDLFEGGTY*

Sequence 1859

Contig_0664_pos_3349_2018,

is similar to (with p-value 6.0e-22)

>gp:gp|AF051917|AF051917_2 Staphylococcus aureus plasmid pSK 41, complete sequence. NID: g3676412.

5 atgaaatattgcagtaattgtggtcaacctcttcgagaaggtgtaaaggtgtgtacaaat
 tgtggtacgcctgtgagaaatgatggacctaattataaacattcagaacaacgttattct
 catcaacaaccacgttccaataagagtaataaaaaaacttgggtgattgttactatagt
 ttagccattatcatttcttggtagtaatctttactatagctaaaaatcaaatgtctcca
 gaaaaacaagcgactcacattgcacatgctatcaaaaaagacgatgctaaatcattatct
 10 aagcaattaacatcaaacgatcatcggttaaatgaagaagaagcacgtgcgtacttaag
 tatattaaagcagaaagtgatttaaaagcatggtgctgacaaagttgaagaaaacaccaa
 gatattaaaaataatcactataacaatttatctgtagatgcaaatgataataatatttta
 aatatatctaaagacgggaaaaaatatgtttttttgataactatcaatttaattgttcct
 caaaaaacaattacgctcgtttcaagtgtatggtgaaattacttatgaatttaacggg
 gataaacatcatatttctgtagaagaagatgacgataaagaattaggaacatttcctatc
 15 ggtgattataattttaaagcatcaaaagacatggaaggtaaaaattttaaaggcgctatt
 acaattgatagtgagtgaaagtgatagttgcataatgaatcggttaaacaaaaacgtttt
 aatggtgatactgaaggcggtatataattagataatgtaaaaatataatgctaattgtaaa
 gaaataggcgatgggttttcatcagaaacatatggctccttatgatccagatgaagaagtt
 atcggtcacgctgaagggttcatacgaaggaaaaacttttaaatcaaatcgggtcaatgta
 20 gcaagtgcagcgaaaaagatgggtggtgtgacagatgtcacagtcgaattcgatgaagaa
 gctattgatcagtatgtcgtataaaaaattagatgaaaaatacagtgattcagatgacgag
 tcagataacgattcaagtagtggtggaagtaacgcgtgaaaatgtaattgataaagtagag
 tcatatgaaggacatacactagatactgatacgtatacgtataaagaacctgaaaaaac
 ggtgatggtaaatggggttttcatcttagataaagatggagatttagctggatcgtag
 25 acggtagacattgcagcaggttatgttacagaatatgacgaagatgggtgaagaagtggga
 tctggttattaa

Sequence 1860

30 MKYCSNCGQPLREGVKVCTNCGTPVRNDGPNYKHSEQRYSHQQPRSNKSNKKTWLIIVTIV
 LAIIIALVVIIFTIAKNQMSPEKQATHIAHAIAIKKDDAKSLSKQLTSNDHRLNEEEARAYLK
 YIKAESDLKHVADKVEENTKDIKNNHYNLSVDANDNNILNISKDGGKYVFFDNYQFNVP
 QKTITLVSSDSGEITYEFNGDKHHISVEEDDDKELGTFPIGDYNLKASKDMEGKNFKGAI
 TIDMSESDSIAYESFKQKRFNVDTEGGYILDNVKIYANGKEIGDGFSSSETYGPYDPDEEV
 IVHAEGSYEGKTFKSNVSVNASASEKDGGVTDVTVKFDEEAIHQYVDKKLDEKYDDSDDE
 35 SDNDSSSGEVTRENVIDKVESYEGHTLDTDTYTYKEPEKTGDGKWGFSFLDKDGLAGSY
 TVDIDDGYVTEYDEDEGEVGSY*

Sequence 1861

Contig_0664_pos_1711_1025,

40 is similar to (with p-value 1.0e-29)

>sp:sp|P44068|Y882_HAEIN HYPOTHETICAL PROTEIN HI0882. >pir:p
 ir|E64015|E64015 hypothetical protein HI0882 - Haemophilus i
 nfluenzae (strain Rd KW20) >gp:gp|U32770|U32770_2 Haemophilu
 s influenzae Rd section 85 of 163 of the complete genome. NI
 45 D: g1573898.

atgaaaaatcaaaaaatcaactatgataaagtattaagaaagataatttctcaatgggaa
 cgtgatggagaacgcctaaaaatcttacttcatagttggtgtgcaccttgtagtacata
 acgttagagtttttaacacaatatgcggatatagcgatttattttgcgaatcctaata
 catcccaaaagtgaatattttaagacgtgctaaagttcaagaacaatttgtaaatgatttt
 50 aataataaaaacaggtgcaaatgtaaagtatattgaagccgaatatgaaccgcataaattt
 atgaaaatggcaaaagataaaggtttaactgaagagccggaaggtggactaagatgtacg
 gcttgtttcgagatgcgattagaaattggttgcaaaagctgctttagaacatggttacgat
 tattttggtagtgcaatcacactctctccaaagaaaaatgcgcaattaatcaatgaacta
 ggtatggatgtacaaaatatataatgtaaaatatttaccagtgatttttaaaaaaat
 55 aaagggtatgaacgttctatcgaaatgtgtaattgattataatatttttagacaatgttat
 tgtggttgtgtatttgcagcgatgaagcaaggtatagatttttaacaaataaataaagat
 gctcaagcattttttacaacaatttttaa

Sequence 1862

MKNQKINYDKVLRKIIISQWERDGERPKILLHSCCAPCSTYTLFLEFLTQYADIAIYFANPNI
HPKSEYLRRRAKVVQEQFVNDNFNNKTGANVKYIEAEYEPHKFMKMAKDKGLTEEPEGGLRCT
ACFEMRLEIVAKAALEHGYDYFGSAITLSPKKNAQLINELGMDVQNIYNVKYLPSPDFKKN
KGYERSIEMCNDYNI FRQCYCGCVFAAMKQGIDFKQINKDAQAFLQQF*

5

Sequence 1863

Contig_0665_pos_3731_4393,

is similar to (with p-value 4.0e-36)

10 >gp:gp|Y17797|EFY17797_7 Enterococcus faecalis gph, ydjH, yd
jG, ydjI, pbp4 and ydiC, ORF2 and ORF3 genes, partial. NID:
g3341430.

15

atgaactatcttctaatgatacgtcaaaccaacctttatcagtagctattatgaaagat
aatgaagtgattgctgaaaaaacaactgatatacaaaaagaatcattcagtgcaactcatg
cccgaatatagcagaaattcttacagaaagtaaaataaaacagaaataactgatatac
gtggtagcgggaaggtccaggttcatataccggtcttagaataggggttactgttgctaaa
acattagcgtatgcattaaacactaacttatatgggtgtctcatcacttaagcacttgct
agcacagtaaaagacagtagcgaagttgttagtagtaccatttttgatgctagaagagaagca
gttttagcaggtgttttatcaatatcaggacaatgaattaataaccattattgatgacact
tatatacctatttttgaacttattgaaaaacttcatcaattaacacacacttatgtgttt
20 gtaggatttcatatcgaaaaataaaacatttattagacagtgacatcgtagaacagtta
ccacaagcttcaagtatgaagcaattaatccaaaaaccagaaaatatacattcatttact
cctaaatatcataaattatcagaggcggaacgaaattgggttaaccaacaagagaacaat
tga

25

Sequence 1864

MNYLLIDTSNQPLSVAIMKDNEVIAEKTDDIKKNHVSQMLPEIAEILTESKINKTEITDI
VVAEGPGSYTGLRIGVTVAKTLAYALNTNLYGVSSLKALASTVKDSTKLLVPIFDARREA
VYAGVYQYQDNELITIIIDTYIPIFELIEKLHQLNQPYVVFVGFHIEKIKHLLSDSDIVEQL
PQASSMKQLIQKPENIHSFTPKYHKLSEAERNWLNQQENN*

30

Sequence 1865

Contig_0665_pos_4405_4827,

putative peptide of unknown function

35 atgagtggtgaagatgttctctaaagtttttgatatagaaagaaatagtttctcacacagt
tcgtggtcaatcgatgcattttatcatgaaatagaaaacaacgaatttgctacatatttt
gttatagaatttagtgacaaaataattggatatgttggtttatggttagtcggtgatcaa
gcacaaattacaacgatagctatatcaaaggcatttagaggctatggtcttgggcaactt
ttacttaaatatgcaatgaactatgcacgtttttcttgatgtgatgagtttagaagta
agaattagataaatgatgttgcaacatgtttataggaatttgggattccaaatgggtgga
40 aaaagaaagaattattatggagaaggcgaggacgcatttagtcatgtgggtgaatttgaaa
tga

Sequence 1866

45 MSVEDVPKVFDIERNFSHSSWSIDAFYHEIENNEFATYFVIEFSDKIIGYVGLWLVVDQ
AQITTIAISKAFRGYGLGQLLLKYAMNYARFSCDVMSLEVRIDNDVAQHVRNLGFQNGG
KRKNYYGEGEDALVMWVNLK*

Sequence 1867

Contig_0665_pos_3707_3267,

50 is similar to (with p-value 1.0e-67)

>gp:gp|U71377|SEU71377_1 Staphylococcus epidermidis autolysi
n AtlE and putative transcriptional regulator AtlR genes, co
mplete cds. NID: g2267238.

55 atgtatcttgggtgtagtactgaaattaaaacatcacaaacttaaaggtaaggatgactac
ttaaatgatataactattaccacccaagcgtaaaaagtattatggaatattcaaatctt
ttacgtaatgatttagatttatctaaaataacaaacaaaacgatttcttagatcaaaga
gtcattaaacgatatggttcactcgtagcccttaacagaattagatgaagacttattgcgt
aagaacaaaagggaatcgactgatagtcagaaagagtcgtattcctcatcacaataaat
gatgaagaagatcaaactaacgaacaaacagacaaaatagcttaaaccggaacgaacag

tacccaaatcaacaagacaacaatcaaaccaatggtgaaaatggtatgataaataatgac
aattatccttacgcacaataa

Sequence 1868

5 MYLGGSTEIKTSQKKGDDYLNIIYYHPSVKSIMYNSNLLRNDLDSKITNKNDFLDQR
VIKRYGSLVPLTEDEDLRKNQKESTDSQKESDSSSQNNDEEDQTNEQTDQNSLNGNEQ
YPNQDNNQTNGENGMINNDNYPYAQ*

Sequence 1869

10 Contig_0665_pos_3219_2749,
is similar to (with p-value 1.0e-88)
>gp:gp|U71377|SEU71377_2 Staphylococcus epidermidis autolysi
n AtlE and putative transcriptional regulator AtlR genes, co
mplete cds. NID: g2267238.
15 atgagtcgtaaaacatatgaaaaaattgctaataattaatggtatgttcaacgtacttgaa
caacaaatcatcatagtaaagatatggcattatttaggaatgaatttttctatgtaaac
catgaacacagagaaaaattatgaagcgcttttgatctactacaaggatagtttagataat
cccgtttagatggcgcatgttacattcttgcattacctgaaatatttaataaggttgat
gtgtttgagtcgtgatttacctttacctgggtttatgatgaaaatggactttctgatata
20 atgaaaagtatcagtggtgcccttcaatatcttattgcagccgctttggaagtgactgat
gtaaatatatttaagccttcaggttttactatggcgatgaacaattggaatattgcccaa
ttacgtattttttggcaatactgcgctatcgtcagaaaagggtttataa

Sequence 1870

25 MSRKTYEKIANINGMFNVLEQQIIHSKDMALFRNEFFYVNEHRENYEALLIYKDSLND
PVVDGACYILALPEIFNKVDVFESDLPTWVYDENGSLDTMKSISVPLQYLIAAALEVTD
VNIFKPSGFTMGMNWNIAQLRIFWQYCAIVRKEAL*

Sequence 1871

30 Contig_0665_pos_2148_1753,
is similar to (with p-value 2.0e-74)
>gp:gp|U71377|SEU71377_3 Staphylococcus epidermidis autolysi
n AtlE and putative transcriptional regulator AtlR genes, co
mplete cds. NID: g2267238.
35 atgctagatgattgctttgaaataagaaagtgtgttttcgtcgaagaacaaggcgtaacca
ctcgaataatgaatttgatcaatatgaagattactcattccatatagtgggatatataaat
ggtgttcctatggcaactgctagaattagacctttaatactcatatttgtaaaattgaa
cgtgtagcaatcatcaagtgtgtatcgtggtccttggtacggtaaaaaatttaatacatgct
attgaaacaattgcaaaaaaacaccaataacaatgaactcactatgaatgctcaattacaa
40 gttcgagacttttacttaaaactaggttactcaccttttggtaaagtattcttagaagaa
aatataaatcatattagtagtaataagttttataa

Sequence 1872

45 MLDDCFEIRKCVFVEEQGVPLENEFDQYEDYSFHIVGYINGVPMATARIRPLNTHICKIE
RVAIIKWYRGLGYGKNLIHAIETIAKKHQYNELTMNAQLQARDFYKLKGYSPFGKVFL
EENINHISMNKFL*

Sequence 1873

50 Contig_0665_pos_0_525,
is similar to (with p-value 0.0e+00)
>gp:gp|U71377|SEU71377_4 Staphylococcus epidermidis autolysi
n AtlE and putative transcriptional regulator AtlR genes, co
mplete cds. NID: g2267238.
55 atgaaagcaccaagaattgaagaagattatacgtcatatttccctaaatatggctataga
aacgggtgtgggacgtcctgaaggtatcgttgttcatgatactgcaaatgataactcaaca
atcgatggcgagattgctttcatgaaacgtaattacacaaatgcattcgtacacgcattt
gttgatggcaatagaaattatagaacagctccgacagattacttatcttggggtgcaggt
ccatatggaaatcaacgttttatcaatgttgaaatcgccatacacatgattatgattca
tttgacagttcaatgaacaactacgctgattatgctgcaacgcaattgcaatattataat

ttaaaacctgatagcgctgaaaacgatggaagaggaacagtttgacacatgctgctatc
tctaacttcttaggaggtactgatcacgctgaccctcaccaatatttaagaagtcacaat
tatagctatgcagaattatatgacttaatttatgaaaaatattta

5 Sequence 1874
MKAPRIEEDYTSYFPKYGYRNGVGRPEGIVVHDTANDNSTIDGEIAFMKRYTNAFVHAF
VDGNRIIETAPT DYLSWGAGPYGNQRFINVEIVHTHDYDSFARSMNNYADYAATQLQYYN
LKPDSENDGRGT VWTAAISNFLGGTDHADPHQYLRSHNYSYAELYDLIYEKYL

10 Sequence 1875
Contig_0667_pos_840_1853,
is similar to (with p-value 0.0e+00)
>sp:sp|P45554|DNAK_STAAU DNAK PROTEIN (HEAT SHOCK PROTEIN 70
) (HSP70). >gp:gp|D30690|STANHS_3 Staphylococcus aureus gene
15 s for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds. NID:
g487326.
atgggtacagattataaagtagatattgaaggtaaatcatatacaccacaagaactttca
gcaatgattttacaaaatttaaaaaagcactgcagaaaactatttaggggatacagtagac
aaagctgttatcactgtccctgcttatttcaatgatgggtgaacgtcaagcaactaaagat
20 gctggtaaaattgcaggcttagaagttgaacgtattatcaacgaacctacagctgctgca
ctgcttatgggttagataaaaactgaaacagatcaaaagggttctcgattttgacttaggt
gggggaacatttgacgtatctattctagagtttaggcgacggcgtatttgaagtattatca
actgccggagataataaacttgggtggcgatgacttcgaccaagtattattgattatctt
gtttcagaattcaagaaagagaatgggtgtagatttatcacaagataaaatggcattacaa
25 agattaaaagatgctgccgaaaaagctaaaaaagatttatcaggtgtttctcaaactcaa
atttcattaccattcatttctgctggagaaaatggcccattacacttagaaattagttta
actcgttctaaatttgagggaattagctgattcattaatcaaaaaactatggaaccgact
cgtcaagcatttaaaagatgctgggttatctacttcagaaatagatgaagttattttagtt
gggtgggtcaacacgtatttccggccgttcaagaagctgttaaaaaagaaattgggaaagaa
30 ccacataaagggtgttaacccagatgaagttgtagcaatgggtgctgctattcaagctggt
gtaatcacaggtgatgttaaagatgtagtattacttgatgttacgccattatcttttaggt
atcgaaattatgggtggacgtatgaacacattaattgaacgtaataactactattccaact
tccaaatcacaagtttattctacagcagctgacaatcaaccagcagtagtgtaa

35 Sequence 1876
MGTDYKVDIEGKSYTPQELSAMILQNLKSTAENYLGDTVDKAVITVPAYFNDGERQATKD
AGKIAGLEVERIINEPTAAALAYGLDKTETDQKVLVFDLGGGTFDVSILELGDGVFEVLS
TAGDNKLGDDFDQVIIIDYLVSEFKKENGVDLSQDKMALQRLKDAAEKAKKDLGVSQTQ
ISLPFISAGENGPHLEISLRSKFELADSLIKKTMEPTRQALKDAGLSTSEIDEVILV
40 GGSTRIPAVQEA VKKEIGKEPHKGVNPDEVVAMGAATQAGVITGDVKDVLVDVTPLSLG
IEIMGGRMNTLIERNTTIPTSKSQVYSTAADNQPAVV*

Sequence 1877
Contig_0667_pos_3958_4263,
45 putative peptide of unknown function
atgctcctttctgctataactccttataaaaaggaggtgaaaatatgaaaagttttattatt
gcgtatgattttaataaccaaaggattatccaaaattaatagacgtattgaggattat
cctaattgttgctaaaaatcaataaatcagtttgggtttattaattcaactaatgatgctaa
actatttagaaacgaattaaaaatgtttattgatagcgatgatagtttgttcgttggttaag
50 ctgactggtgaagccgcatggtctaattgtaatttgcagttcacacattttaaagattat
cttttag

Sequence 1878
MLLSAILLIKGGENMKSFIIAYDLNNQKDYPKLIERIEDYPNVAKINKSVWFINSTNDK
55 TIRNELKMFIDSDSLFVGKLTGEAAWSNVICSSQHLKDYL*

Sequence 1879
Contig_0667_pos_4836_5150,
putative peptide of unknown function

atgtgctttttcaaaaagaatgaaacaatcaagagaaaaacaaggatgacttttagctgaa
ctaggaagaaaaatcggtaaaactgaagctactgtacaacggtatgaaagcgggaatatt
aaaaatcttaaaaatgatactattgaaagtatagctactgcattaaatgttaaccctgct
5 ttcttgatgggttgatagaagaagttgaggaacaaccacaacatcgtgcagcgcacatt
gatgggtgatttaactgacgaagaatggcaagaattcttgattacgctgaatacataaga
agtaaaaagaaaataa

Sequence 1880

MCFSKRMKQSREKQGMTLAE LGRKIGKTEATVQRYESGNIKNLKNDTIESIATALNVNPA
10 FLMGWIEEVVEEQPHRAAHL DGDLTDEEWQEILDYAEYIRSKRK*

Sequence 1881

Contig_0667_pos_5154_5624,
putative peptide of unknown function
15 gtgttttatgtggggaaatatgaagatatgttaattgaacatgactatattgaagtcatt
gaatgtgataacttacctaaaagggttatctggtttgtggcttgagatatgattttaatt
aatcgtaacttgccattacttccaaacttgaaacacttgacaggaactcgctcataac
gaacttacatatggaaatatagttgatcaaagtagttttaatcatagaaaattgaaggt
tatgcacgtaggttagcctatgaaaagttatccctcttaagatatgtaaaagcattt
20 ttgcaaggcattcatgacttgatgaacttgctaatttttttgaggttacagaagggtttt
gtcctacaaagtattgaacattataaacaataatgggtattccactcggtatagcaaa
tacgttattcaatttgagccgttacgagtggttgatatataagatatatag

Sequence 1882

25 VFYVGKIEDMLIEHDYIEVIECDNLPKRSLGLWLGDMILINRNLPTSKLET LAEELAHN
ELTYGNIVDQSSFNHRKFEGYARRLAYEKLIPKLDIVKAFLQGIHDLYELANFFEVTEGF
VLQSIHYKQKYGYSTRYSKYVVIQFEPLRVFEYKDI*

Sequence 1883

30 Contig_0670_pos_3246_2653,
putative peptide of unknown function
atgggccctcaatattggtggccagcagaaacgccaatagaaatgatgcttggggcaatt
ctagtcctaaaataactaattggaacaatgcagatatagcggtatcaagattaaaagaagaa
acttcttttaattgcacagacgatatgtgaaaatgcctttagaatcggtgcagcaagtgata
35 cgttcgagtggtttctataaaaaataaagctaaggctatacaggcattgttactatggtta
aatcaacatcatttttgattatagtagtatagctaagttatacggtgatagcttaagaaaa
gaattactcaccatccgtggtataggtgaagagaccgccgatgtcttaatagtatatatt
tttaaaaggtaaagaattcatacctgatagttatactagacgtatttttagaaaattggga
tatcaacatacagaaagttatcataaattgaaacaggaattaacacttcctgaatcattt
40 tcaaatcaagatgcaaatgagtttcacgctttattagataattttgggaaaaattatttt
aatggtaaggggaacaacgcgtataccttttttagatacctattttaaaaaataa

Sequence 1884

45 MGPQYWWPAETPIEMMLGAILVQNTNWNADIALSRLKEETSFNAQTILKMPLES LQQVI
RSSGFYKNKAKAIQALLWLNQHHFDYSSIAKLYGDSLRLKELLTIRGIGEETADVLIVYI
FKGKEFIPDSYTRRIFRKLGYQHTESYHKLKQELTLPESFSNQDANEFHALLDNFGKNYF
NGKGKQRYTFLDTYFKK*

Sequence 1885

50 Contig_0670_pos_2244_1846,
putative peptide of unknown function
atgcaaaattttcaatttaaatgaacagtatgaaaaagaagcagcaagtaagtatggagat
actcattattatcaagcatataaagataaacaataatgtaaggatgaatcagaacagcaa
aatcatttttgaggaaattaataagcaattaaatatgttttttgacgaaatgaatcaacta
55 tacttaaacaaagtttctatacttgaaagcaagtggaataactaagaaattacaatgtatt
ttgaaagaacaagttcctaattgtgacaatcaatttttagaataatagctcagattttat
attgaggacgagcgattttgtaaagttttattaacaagcaacgtgaacgtggtttgaattta
tacatttctgacgcgataaaaaacattttattaaattataa

Sequence 1886

MQNFNLNQEYKEAASKYGDTHYYQAYKDKQCKKDESEQQNHFEINKQLNMFFDEMNL
YLNKVSILEASGKTKKLQCILKEQVPNCNDFLEYIAQIYIEDERFVKFINKQRERGLNL
YISDAIKTFIKL*

5

Sequence 1887

Contig_0670_pos_1409_510,

putative peptide of unknown function

gtgatgctccaaatgaaaagtaaactatggatagttatatgtgcattgattgtagtactg
10 gctgcatgtggtcaagatgccaatcattcatctaataataaagacactgaaaaagcgat
aaaaaatatcatagaattatttcgctcattcctagtaacacagaaattttatatcgctta
ggaatcggagaagatatagttggtgtatccactgtggatgattatcctaaagatgtaaaa
aaaggtaaaaaacaattcgatgcatgaatttaataaagaagaattaataaaagctaaa
ccggatttgatttttagcgcatgagtcacagaaaaattctgcaggtaaagtgtctaaagtca
15 cttaaagataagggagtaaaaagtcgtttatgtgaaagatgcacaatcgattgatgaaact
tatgatacttttaaatcaattggacaattaacggatcggtgaaaaacaagctaaagaactt
gttgatgaaacaaaacacaatgtagaaaaaatcattaactccgttcctaaacatcataag
aaacaagaagtgttttatggaagtatcgctctaaaccagacatttacactgccggaaaagat
accttcttttaacgatatgttagagaaaactagatgctaaaaatagttttgatgatgttaaa
20 ggttggaatcagtaagtaaaagaaagcattattaaacgtaatcctgatattctgatttcc
acagaaggtaaaatcaaaatcagactacatagaaatgataaaaaaacgtggcggttttgat
aaaattaatgctgttaaaaatcacgtattgaaacagtagatggggatgaagtttctcga
ccaggtcctcgattgatgaaggtctaaaggatttaagagacgatatatataaaaaatag

25

Sequence 1888

VMLQMKSKLWIVICALIVVLAACGQDANHSSNNKDTEKSDKKYHRIISLIPSNTEILYRL
GIGEDIVGVSTVDDYPKDVKKGKQFDAMNLNKEELIKAKPDLILAHESQKNSAGKVLKS
LKDKGVKVVYVKDAQSIDETYDTFKSIGQLTDREKQAKELVDETKHNVEKIINSVPKHHK
30 KQEVFMEVSSKPDIIYTAGKDTFFNDMLEKLDKNSFDDVKGWKSVSKESI IKRNPDIILIS
TEGKSKSDYIEMIKRGGFDKINAVKNTRIETVDGDEVSRPGPRIDEGLKDLRDDIYKK*

Sequence 1889

35 Contig_0673_pos_1644_1195,

putative peptide of unknown function

gtgctaaagcaaacggttgattatagaaattgcaagatttgttcctagtagtgaacttaaa
aataaaaatatataaaaagcttttaaaaatggatgtcgggaatcatacttcatttgcatat
aaagtgttcctgattgttttatccagaatacatttcggttggaagaatacagtcatt
40 ggttataatacaacaatattaactcacgaagtacttggtgatgagtgagagtaggaaaa
gtcattataggcgattacactttaataggtgcaaatacgacaataattaccaggaataacc
ataggaaatcatgttaaaattggtgcgggtacggttggtgtctaaagatgttcccgattac
agttttgcatttggtaatcctatgcaaatacaattagattcaggaggtgacaatgaatgg
cacaaaaagaaaataacatcattccaatga

45

Sequence 1890

VLKQTLIIETARFVPSMKLNKIYKLLKMDVGNHTSFAYKVLPLDFYPEYISVGKNTVI
GYNTTILTHEVLVDEWRVGKVIIGDYTLIGANTTILPGITIGNHVKIGAGTVVSKDVPDY
SFAFGNPMQIQLDSGGDNEWHKKKITSFQ*

50

Sequence 1891

Contig_0673_pos_0_868,

putative peptide of unknown function

atgcttgaaaaaacattcgaagtcacgtatacaaatgaacaaaaaattgaattagaagca
55 caattgttttcaacacaacttttatttcaatttctcttttcgcaaggtagggttagaagaa
gcccgaaatataattttgaatcaatcttacgagatacaacagcatagggtgattaggaat
ttacttgcaatgtgttatttctatctaggtgagtagatgtagcgccaaagcaatgtttgaa
gaacttttaaggaagataattcagacgtgcatgcactttgtcactacacattattactt
tataataaaaaagaaacagaaaaatatcaaaaatatcttaaaatacttaataaagtagta

ccactaaatgacgacgaaacctttaattaggaatcgtattgagttattttaaacagtat
 cgtgcttctcaaaatttactttatccactttataaaaaaggtaaatttgtctctattcaa
 atgtataatgcattgagtttcaatttttattacctaggaataaaagacgaaagtattgag
 atgtggaacaagctcactcaaatttctgaagttgatgttggttatgcaccttgggtaatt
 5 gagaaaagtaaaacggtattttgaatcacgagtggtaccattattactagatgataataat
 cattatcgactttacggtatttttttacttcatcaattaaatggaaaagaaataactaatg
 actgaagatatttgggtcaattcttgaatcaatgaatgactatgagaaactttatctcaca
 tatttggtagaaggactcacactcaataaattagattttatacacagaggtatgcaaagg
 ttgtataattttaagaaattcCAAGAAA

10

Sequence 1892

MLEKTFEVITYTNEQKIELEAQLFSTQLLFQFLFSQGRLEEARTYILNOSYEQQHRVIRN
 LLAMCYLYLGEYDSAKAMFEELLKEDNSDVHALCHYTLNKKETEKYQKYLKILNKVV
 PLNDDFTFKLGIVLSYKQYRASQNLPLYKKGKFVSIQMYNALSFNFYLLGNKDESIE
 15 MWNKLTQISEVDVGYAPWVIEESKTVFESRVLPLLLDDNNHYRLYGIFLLHQLNGKEILM
 TEDIWSILESMDYKLYLTYLQGLTLNKLDFIHRGMQRLYNFKKFQEX

Sequence 1893

Contig_0675_pos_348_1073,

20 is similar to (with p-value 2.0e-48)

>gp:gp|AF080002|AF080002_2 Heliobacillus mobilis exopolyphos
 phatase Ppx (ppx) gene, partial cds; cobyric acid synthase C
 obQ (cobQ), UDP-N-acetylmuramyl tripeptide synthetase MurC (
 murC), glutamyl tRNA reductase HemA (hemA), photosynthesis g
 ene cluster, complete sequence, stage II sporulation protein
 25 E Sp2E (sp2E), cell cycle protein MesJ (mesJ), and ATP-depe
 ndent zinc metalloproteinase FtsH (ftsH) genes, complete cds;
 and nucleoside diphosphate kinase B NdkB (ndkB) gene, parti
 al cds. NID: g3820536. >gp:gp|AF080002|AF080002_2 Heliobacil
 30 lus mobilis exopolyphosphatase Ppx (ppx) gene, partial cds;
 cobyric acid synthase CobQ (cobQ), UDP-N-acetylmuramyl tripe
 ptide synthetase MurC (murC), glutamyl tRNA reductase HemA (
 hemA), photosynthesis gene cluster, complete sequence, stage
 II sporulation protein E Sp2E (sp2E), cell cycle protein Me
 35 sJ (mesJ), and ATP-dependent zinc metalloproteinase FtsH (fts
 H) genes, complete cds; and nucleoside diphosphate kinase B
 NdkB (ndkB) gene, partial cds. NID: g3820536.

atgaatgaactaacggtttatcatttcatgtcagataagcttaattttatacagtgatatt
 ggtaatatcatggcattaaacaaagagctaaaaaagaaatattaagataaatgttaaa
 40 gaaatcaatgagactgagggagtcacatttgatgattgtgatattttcttcattgggtggt
 gggagtgatagggaacaagcgcttgccacgaaagaattaagtaaaattaaaacttcttta
 aaaaatgcaattgaagatggtatgcctgggttaactatatgcggtgggttatcaattttta
 ggtcataagtatattactcctgatggtaccgagttagaaggattgggtggttcttgacttc
 tataccgagtcataaaaaagaacgcttaactggagatatcattatagagagtgatactttt
 45 ggcacgattgttgatttgaaaatcatggtggttagaacatatcatccgtatggaacatta
 ggccgagtaacgtatggttatggttaataatgataacgatcgaaaagaaggtatacactat
 aaaaatctattaggttcttatcttcacggtccaattttacaaaaaatcatgaaataact
 gattatctacttgagaaagcatgtgaaagaaaaggatactatttgagcctaagaagatc
 gataacacagaggaagaagctgctaagcaagttctgattaacgtgcaaaagaaaataaa
 50 aaataa

Sequence 1894

MNELTVYHFMSDKLNLYSIDIGNIMALKQRAKKRNIKINVKEINETEGVTFDDCDIFFIGG
 GSDREQALATKELSKIKTSLKNAIEDGMPGLTICGGYQFLGHKYITPDGTELEGLGVLD
 55 YTESKKERLTGDIIESDTFGTIVGFENHGGRTYHPYGLGRVITYGYGNNDNRKEGIHY
 KNLLGSYLGHPILPKNHEITDYLLEKACERKILFEPKKIDNTEEEAAKQVLIKRAKENK
 K*

Sequence 1895

Contig_0675_pos_3361_4116,
is similar to (with p-value 2.0e-44)
>sp:sp|P19994|AMPM_BACSU METHIONINE AMINOPEPTIDASE (EC 3.4.1
1.18) (MAP) (PEPTIDASE M). >pir:pir|JS0493|JS0493 methionyl
5 aminopeptidase (EC 3.4.11.18) - Bacillus subtilis >gp:gp|L47
971|BACRPLP_16 Bacillus subtilis ribosomal protein (rplPNXEF
ROQ, rpmCDJ, rpsQNHEMK) genes, integral membrane protein (se
cY) gene, adenylate kinase (adk) gene, methionine aminopepti
dase (map) gene, initiation factor 1 (infA) gene, RNA polym
10 erase alpha (rpoA) gene. NID: g1044970. >gp:gp|Z99104|BSUR00
01_138 Bacillus subtilis complete genome (section 1 of 21):
from 1 to 213080. NID: g2632267. >gp:gp|D00619|BACSECY_5 Bac
illus subtilis genes for ribosomal proteins, SecY, adenylate
kinase and methionine amino peptidase, complete cds. NID: g
15 216336.
atgattgttaaaactgatgaagaattacaagcgtaaagaagaatagggtacatttgtgca
aaagtcagagatactatgaaagaagctactaaacctggagtgactaccggtgaattagat
catattgccaaagatttatttgaagagcatggggcgatatcagcgctattcacgatgaa
aacttcccaggtcaaacttgcatttagtgtaacgaagaggtcgcgcatggaatccctggt
20 aacgagtaattcgtgaaggtagcttagttaatatgtatgtatcagctttaaaaaatggg
tactatgctgacactggaatttcatttggtagggaaatcagatcaaccacttaacaa
aaggtttgtgacgtagccacaatggcttttgaagaatgctatgaaaaaggtgaaaccaggt
acaaaattaagtaatttggaaaagctgttcatgcaactgcacgcaaaatgacttgact
gtgattaaaaatttaactggacatggtgttggacaatcacttcatgaggcacctaatac
25 gtcattgaatttttgcctaaagataaaacattattaaaagaagggaagtcattgca
gtagaaccattcatatcaacacatgctacatttgaactgaaggtaaaaatgaatgggct
tttgaactaaagataaaaagttatgtcgctcaaattgaacacacgggttagttacaaaa
gatggtcggttacttacaactaagattgatgattaa

30 Sequence 1896
MIVKTDEELQALKEIGYICAKVRDTMKEATKPGVTTRELDHIAKDLFEEHGAISAPIHDE
NFPQGQTCISVNEEVAHGIPGKRVIREGDLVNIDVSALKNGYYADTGISFVVGKSDQPLKQ
KVCADVATMAFNAMKKVPGTKLSNIGKAVHATARQNDLTVIKNLTGHGVGQSLHEAPNH
VMNYFDPKDKTLLKEGQVIAVEPFISTHATFVTEGKNEWAFETKDKSYVAQIEHTVIVTK
35 DGPLLTTKIDD*

Sequence 1897
Contig_0675_pos_4702_0,
putative peptide of unknown function
40 atgttgattatttttactgcttttaattgattatttactatattttttgaa
aaaattggctttttactagtaactcctattaggatgtgtacttgtatatgtagggtatgtg
tattttcataaagtaagaggactactatctttttggataggaaccttattaattgctttt
acacttttgcctaataagtaacacgataattattctatttataatttttaatagtagtcac
atacgttatttgggtttataagtttagacctttaaagtgattgctacagatgaagaaatc
45 acatcacccatttttattaagcaaaaatggtttgggaacaacatacaccagtgatgta
tataaatgggaagacgtacagattcaacacgggtataggagacatacacattgatatgaca
aaagcggcaaatattaaggaacaaataccatagttgtgctcatatttttaggtaaaagta
caagtagttgtacctcttaattataataaaatttcatgacgactctcttctacggcact
gcttatgtgaacgataaatcttataagattgagaataaccatgttcaaattgaagaaaaa
50 acgaaagatgataattatactgttaattgtttacgtttcat

Sequence 1898
MLIIFTALMIANFYIIFFEKIGFLLVLLGCVLVYVGYVYFHKVRGLLSFWIGTLLIAF
TLLSNKYTIIILFIFLIVVIRLVYKFRPLKVIATDEEITSPIFIKQKWFGEQHTFVYV
55 YKWEDVQIQHGIDIHIDMTKAANIKETNTIVVRHILGKVQVVVPLNYNINLHATLIFYGT
AYVNDKSYKIENNHVQIEEKTDDNYTVNVVYSX

Sequence 1899
Contig_0675_pos_3106_2204,

putative peptide of unknown function

5 atgttggttaaaccctcactccaatttttgccatacttactgcaattgttacaattgaacct
actgctaaagcatcattaaaaaagggtataaaaaggctgccagcaacagttatcggtgcg
ttatttgctgttgcctttacatatgtcttcgggtgatcaatcaccggttaagttatgcttta
10 agtgcacattttaccattctgatatgcactaaaacttaatttacaggttaggaacaactgtc
gcagtattaaacttccggttgcaatgattccagggtatacatgaagcatatgtgttcaatttc
ttttcacgggttacttacagctcttataggacttggtacagctggattagtcaattttatc
atcttaccacctaagattatcatcaacttgaagagcaattagcccttagtgagaaaaaa
atgtatcgtttattttatgaacgctgtaatgagttattattaggaaaattcagctcggaa
15 aagactagtaaagaattatcaaaatataatatttgcctcaaaaagttgaaacattaatg
agttaccaaaagagatgaacttcattatcataaaaaatgaagataattggaaattattaaat
cgcttatacaaatcgcgcttataacaacggtttatttatttcacatttatctaactattt
tatttaccacaacatacgtctattgcttttgatgctaatagagaagatagcattgattaat
attagtaatagtattaatgacatcattcaaaaaggaagctttgcacgtcaaaaaaaatct
15 attgcaacactaaagtcttctgttaaacagatggatgagtttgaccaaatacaaatgaaa
agtacactcatatatgaaattctactcatatacaaaaatttttagattcacggttatgcaaaa
taa

Sequence 1900

20 MLLNLTPIFAILTAIVTIEPTAKASLKKGYKRLPATVIGALFAVVFTYVFGDQSPLSYAL
SATFTILICTKLNQVGTTVAVLTSVAMIPGIHEAYVFNFFSRLLTALIGLVTAGLVNFI
ILPPKYYHQLEEQALALSEKKMYRLFYERCNELLLGKFSSEKTSKELSKLNIJAQKVETLM
SYQRDELHYHKNEEDNWKLLNRLTNRAYNNRLFISHLSNIIYLPKHTSIAFDANEKIALIN
25 ISNSINDIIQKGSFARQKKSIA TLKSSVKQMDEFDQNMKSTLIYEILLIYKILDSRYAK
*

Sequence 1901

Contig_0678_pos_933_2351,

putative peptide of unknown function

30 gtgattgaattaattaaaatggaagggatgatagttgtgtcctaataataatttttaaagat
gatttcgaaaagaatcgtcaatctattaatccagacgaacatcaaacagaattaaaagaa
gatgataaaacaaatgaaaataaaaaagaagctgactctcaaacagtttatctaataac
tcaaatcaacaatttcctccgagaaatgcccaacgacgaaaaagacgcagagagacagca
actaatcaaaacaaacaacagcagcaaacatcaaaaaaatagtgcgctaaaactaca
35 gaaggttcattagatgaccgttatgacgaagcacagttacagcaacaacatgataaatcg
caacaacaaaataaaactgaaaaacaatcacaagataatagaatgaaagatggaaaagat
gcagctattgtaaatggaacatctgagtcaccagaacataaatcaaaatcaacacaaaat
agacccggccctaaagctcaacaacaaaagcgtaaatcagaaagtacgcaatcaaaaccg
tcaacaacaaagataaaaaagcagctacaggtgctggaatagctgggtgcagctgggtgt
40 tctgggtgcagcagaacatccaaacgctcatcataataaaaaagataaacaagattctaaa
cactcaaacatgagaatgacgaaaaatctgttaaaaatgatgaccaaagcaatctaaa
aaaggcaaaaagcagcagtcggtgctggcgagctgcaggagttgggtgcggtggtgtt
gcgcatcataataatcaaaataaacatcataatgaggaaaaaaattctaatacaaacat
cagtacaatgaccaatcagaaggtaagaaaaaagggtggtttcatgaaaatcttgttacca
45 cttatagcagccattcttattctaggtgcaatagcaatattcgggtggtatggctctaaat
aatcacaacgatagtaaaaagtgatgacaaaaaatagcgaatcaaagtaagaaagactca
gataaaaaagatgggtgcgcaatccgaagataacaaagacaaaaaatctgatagtaacaaa
gacaaaaaatctgattctgataagaacgcagatgatgactctgataatagttcctcaaat
cctaacgctacttcaactaataaacgataatgtagccaataataactcaaattatataca
50 aacaaaaatcaacaagataatgcaaaccaaaatagcaataatcaacaggcaactcaaggt
caacaatcacatacagtatacgggtcaagaaaaacttatatcgatcgccatacaatattat
ggagaaggaaactcaagctaacgtagataaaattaaacgtgcgaatggattaagcagtaat
aatattcataatgggtcaaacattagttattcctcaataa

55 Sequence 1902

VIELIKMEGMIVVSNNNFKDDFEKNRQSI NPDEHQTELKEDDKTNENKKEADSQNSLSNN
SNQQFPFPRNAQRRKRRTATNQSKQQDDKHQKNSDAKTTEGSLDDRYDEAQLQQQHDKS
QQQNKTEKQSQDNRMKDGKDAI VNGTSESPEHKSSTQNRPGPKAQQQRKSESTQSKP
STNKDKKAATGAGIAGAAGVAGAAETSKRHHNKKDKQDSKHSNHNDEKSVKNDDQKQSK

KGKKAAVGAGAAAGVGAAGVAHHNNQNKHHNEEKNSNQNNQYNDQSEGKKKGGFMKILLP
LIAAILILGAIAIFGGMALNNHNSKSDQKIANQSKKSDKDKDGAQSEDNKDKKSDSNK
DKKSDSDKNADDDSDNSSSNPNATSTNNNDNVANNNSNYTNQNNQDDNANQNSNNQATQG
QQSHTVYQGQENLYRIAIQYYGEGTQANVDKIKRANGLSSNNIHNGQTLVIPQ*

5

Sequence 1903

Contig_0678_pos_2484_3470,

is similar to (with p-value 0.0e+00)

>sp:sp|P50736|YPDA_BACSU HYPOTHETICAL 36.3 KD PROTEIN IN REC
10 Q-CMK INTERGENIC REGION. >gp:gp|Z99115|BSUB0012_235 Bacillus
subtilis complete genome (section 12 of 21): from 2195541 t
o 2409220. NID: g2634478. >gp:gp|Z99116|BSUB0013_7 Bacillus
subtilis complete genome (section 13 of 21): from 2395261 to
2613730. NID: g2634723. >gp:gp|L47648|BACSERA_12 Bacillus s
15 ubtilis phosphoglycerate dehydrogenase (serA), ypaA, ferredo
xin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutama
te dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex lyti
c enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate ki
nase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent gly
20 cerol-3-phosphate dehydrogenase (glyc), yphE and yphF genes,
complete cds. NID: g1146195. >gp:gp|L47648|BACSERA_12 Bacil
lus subtilis phosphoglycerate dehydrogenase (serA), ypaA, fe
rredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, gl
utamate dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex
25 lytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophospha
te kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependen
t glycerol-3-phosphate dehydrogenase (glyc), yphE and yphF g
enes, complete cds. NID: g1146195.
atgcaaaactattgaaagcatcattataggtggcggtccatgcggattaagcgcggaatt
30 gaacaaaaagaaaaaggtattgaaacattagtaattgaaaaaggtaatgttggtgaatca
atctataattatccaacacatcagacttttttctcatcaagtataaataagtatcggc
gatattcctttttattgttgaaagtagtaagccaagacgtaatacaagcgcttgatatattat
aggggaagtcggttaaacaatcattcaacttaacatacatccattcgaagaagttttaacagtt
aaaaaaataaacaataaatttgcaattacaactacaaaaggtgtatatgaatgtaaatat
35 ttaactgttgctacgggttattatggtcaacataaacactttagaagcggaaggggcagaa
ttacaaaaagttattccattacttttaagaagcacatccgtatttttaatacaaatgttggt
attattggaggcaaaaactctgctgttgatgctgccttagaattagaaaaagctggtgct
aatgtaactgttttatatcgtggcggaacagtagcctaaagcaattaaacatggatatta
cccaatttcgaatcattagtcacacgaaaaaattacgatggaatttaatgcgacagta
40 accaaaattaccgatcattcagtgacttatgaaaagatgggtcaacttatagaaattgat
aatgactacggtttttgctatgattggttatcatccagattacgatttcttaaaaaacaata
ggtattgatatccataccaatgaatatggaactgctcctgtttataatcgagaaacattc
gaaacaaacgctcgaataattgttatatagctggtgttattgctgcgggtaattgatgcaaat
actatttttatcgaaaatggttaaatatcatggtggtgtcattacacaaagcattttgaca
45 aaaaaacaaacacctcttgaacatag

Sequence 1904

MQTIESIIIGGGPCGLSAAIEQKKKGIETLVIEKGNVVEIYNYPTHQTFSSSDKLSIG
DIPFIVEDSKPRRNQALVYYREVVKHHQLNIHPFEEVLTVKKINNKFATITTKGVYECKY
50 LTVATGYYGQHNTELEAGELPKVFHYFKEAHPYFNQNVVIIGGKNSAVDALELEKAGA
NVTVLRYGEQYPKAIKPWILPNFESLVNHEKITMEFNATVTKITDHSVTYEKDGQLIEID
NDYVFAMIGYHPDYDFLKTIGIDIHTNEYGTAPVYNRETFETNVENCYIAGVIAAGNDAN
TIFIENGKYHGGVITQSILTKKQTPLET*

55 Sequence 1905

Contig_0678_pos_4729_5376,

is similar to (with p-value 4.0e-52)

>sp:sp|P38493|KCY_BACSU PROBABLE CYTIDYLATE KINASE (EC 2.7.4
.14) (CK) (CYTIDINE MONOPHOSPHATE KINASE) (CMP KINASE). >gp:

gp|U11687|BSU11687_4 Bacillus subtilis 168 jofA, jofB, MssA homolog (jofC) and ribosomal protein S1 homolog (jofD) genes, complete cds, and joeB gene, partial cds. NID: g533101. >g
 5 p:gp|Z99115|BSUB0012_229 Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220. NID: g2634478. >
 gp:gp|Z99116|BSUB0013_1 Bacillus subtilis complete genome (section 13 of 21): from 2395261 to 2613730. NID: g2634723. >g
 10 p:gp|L47648|BACSERA_19 Bacillus subtilis phosphoglycerate dehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent glycerol-3-phosphate dehydrogenase (glyc), yphE and yphF genes, complete cds. NID: g114619
 15 5. >gp:gp|L47648|BACSERA_19 Bacillus subtilis phosphoglycerate dehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent glycerol-3-phosphate dehydrogenase (glyc), yphE and yphF genes, complete cds. NID: g1146195.
 20 atgagttcaattaatattgcactagatggcccagctgctgcaggttaagagtacaattgctaaacgtgtagccagtcgtctatcaatgatatatgttgatacaggagcaatgtatcgtgccattacatataaatatttacaataatggcaaacccgaaaattttgattatctgattaataacactaaacttgagcttacttatgatgaagtaaaaggcgaagaatcttactagataatcaagacgtcactgattatttaagagaaaatgatgtaacacatcacgtatcttatgttgcattctaaagaaccagtgctgttcatttgcagtgaaaatacaaaaagaattagctgctaaaaaagggtatcgttatggatggccgagatatttggtacagttgtattaccagatgccgaattaaaagtt
 30 tatatgattgcatctgttgcgtgaacgtgctgaacgtcgacaaaaagagaatgagcaacgtggcattgaatcaaattttagaacaattaaaggaggaaattgaagcacgagatcattatgatatgaatcgtgaaatttcgccattacaaaaagccgaagatgctattacacttgatacaactggcaaatctatagaagaggtaacaaatgaaatattatctctactttaa

35 Sequence 1906
 MSSINIALDGPAAAGKSTIAKRVASRLSMIYVDTGAMYRAITYKYLQNGKPFNFYDLINN
 TKLELYDEVKQGRILLDNQDVTDYLRENDVTHHVSIVASKEPVRSFAVKIQKELAAKKG
 IVMDGRDITGVVLPDAELKVYMIASVAERAERQKENEQRGIESNLEQLKEEIEARDHYD
 MNREISPLQAEDAITLDTTGKSIEEVTNEILSLL*

40 Sequence 1907
 Contig_0678_pos_5860_7038,
 is similar to (with p-value 1.0e-92)
 >sp:sp|P38494|RS1H_BACSU_30S_RIBOSOMAL_PROTEIN_S1_HOMOLOG. >
 45 gp:gp|U11687|BSU11687_5 Bacillus subtilis 168 jofA, jofB, MssA homolog (jofC) and ribosomal protein S1 homolog (jofD) genes, complete cds, and joeB gene, partial cds. NID: g533101.
 >gp:gp|Z99115|BSUB0012_228 Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220. NID: g2634478
 50 . >gp:gp|L47648|BACSERA_20 Bacillus subtilis phosphoglycerate dehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (ypcA), ypdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD+ dependent glycerol-3-phosphate dehydrogenase (glyc), yphE and yphF genes, complete cds. NID: g1146195. >gp:gp|L47648|BACSERA_20 Bacillus subtilis phosphoglycerate dehydrogenase (serA), ypaA, ferredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH, glutamate dehydrogenase (

ypcA), ypdA, ypdB, ypdC, spore cortex lytic enzyme (sleB), ypeB, ypfA, ypfB, cytidine monophosphate kinase (cmk), ypfD, ypgA, yphA, yphB, yphC, NAD⁺ dependent glycerol-3-phosphate dehydrogenase (glyc); yphE and yphF genes, complete cds. NID : g1146195.

5 atgactgaagaattcaatgaatcaatgattaatgatattaaagaaggtgacaaagtcact
gttgaagttcaacaagtagaggataaacaagttgtgtgcatattaatggtggcaaattt
aatggaattattcctattagccagctttcaacacatcatatcgaaaaccctagtgaagtt
gtaaaagtcggtgatgaagtcgaagcatatgtcactaaaatcgagttcgacgaagaaaat
10 gatactggggcatacattttatcaaaaagacaacttgaaaactgaaaaatcctatgaatat
ttacaagaaaaactagataacgatgaagtgattgaagctgaagttactgaagtagttaa
ggtggttttagtcggtgacgttggtcaaagagggtttgtacctgcttctcctaatttcaact
gatttcattgaagatttttctgtattcgatggtcaaacaatccgtattaaagtggaaaga
cttgatcctgaaaaacaatagagtcatttttaagccgtaaagctgtggaacagttagaaaac
15 gacgctaaaaaagcttcaatattagattcttttaaatgaaggcgatgttatgatggtaaa
gttgctcgattaaactaacttttggtgctttcattgatattggtggcgtagatgggttagtt
cacgtttctgaattatctcatgaacatgttcaaacaccagaagaagttgtgtcagtaggt
gaagcagtcaaaagttaaagttaaatctgtagaaaaagattctgaacgtatttctttatct
attaaagacacttttaaacaccatttgaaaaacattaaagggaatttcacgaagatgat
20 ttattgaaggtactgtagtacgtttggcgaaactttggcgcatctcgtagaaattgctcca
tcggtccaaggttttagtgcattttctgaaatcgatcataaacatatcggttctcctaac
gaagtattagaacctggacaacaagttaatgtaaaaatattaggtatcgatgaagataat
gaaagaatttcattatcaatcaaagcaacgttacctaagaaaatgtcattgaaagtgc
gcatccacaactcaatcatatcttgaagatgataatgatgaagataaaccaacattaggc
25 gatgtttttggtgataaatttaaagaccttaagtttta

Sequence 1908

MTEEFNISMINDIKEGDKVTVEVQQVEDKQVVVHNGGKFNGIIPISQLSTHHIENPSEV
VKVGDEVEAYVTKIEFDEENDTGAYILSKRQLETEKSYEYLQEKLDNDEVIEAEVTEVVK
30 GGLVVDVQGRGFVPASLISTDFIEDFSVFDGQITIRIKVEELDPENNRVILSRKAVEQLEN
DAKKASILDSLNEGVDIDGKVARLTNFGAFIDIGVDGLVHVSELSHEHVQTPEEVVSVG
EAVKVKVKSVEKDSERISLSIKDTLPTPFENIKGKFHEDDVIEGTVVRANFGAFVEIAP
SVQGLVHISEIDHKHIGSPNEVLEPGQQVNVKILGIDEDNERISLSIKATLPKENVIESD
ASTTQSYLEDDNDEKPTLGDVFGDKFKDLKF*

35

Sequence 1909

Contig_0678_pos_7112_8380,
is similar to (with p-value 0.0e+00)
>gp|D21131|STASRM551A_1 Staphylococcus aureus gene for a
40 participant in homogeneous expression of high-level methicil
lin resistance, complete cds. NID: g531264.
gtgataattggagggttcaagtttagattcatctcaattattacaagctctatacgaaaca
ttgtatatggtgactgtatcacttgtaatcggtgctttaataggtatacctcttggtcatc
ttgttagtggttaactagaaaaaacggtatatggtcgaatacaatattgcatcaagtaatta
45 aatcctatcattaatatatttaagatcaattccgttcattattttattaatagccatagtg
ccttttattattattgtaatatcaaaaaaattagatttagtagatcgctcctaatttcaga
aaagtacatacgaaacctatctcagtgatgggagggaacggtcattttattttcttctta
atagggatttggctcgacaccctattgaacgtgaggttaaaccgcttatattaggtgca
attacaatgtatatggttggtgattgattgatgatatttacgatctaagaccttatttaaag
50 ttagcagggtcaaattgttgacgctttaattgttacggttttatggaattacaatagacttt
atttcattgccaattggtccaacgattcattttggcatattcagattcctattacagta
atatggattgtagcaattaccaatgctattaatcttatcgacggacttgatggacttgcc
tcaggcgctcagcattggcattaatgactattggattcatcgctattttacaagcgaaac
atatttattatcatgatttgctgtgtacttttagggctttacttggtttcttattctat
aactttcaccagcgaaaaattttcctaggtgatagtggtgcattaatgataggatttatt
55 atcgggtttcttactccttactcggttttaagaatatcacatttattgcattattcttctcct
atagttatattagcgggtgccatttattgatacattatttgcaatgattcgtcgaatgaaa
aaagggcaacatataatgcaagcgggacaagtcacatttacatcataaattacttgcttta
ggatatacgcatagacaaaccgttttacttattttattcaatagcgattatgttttagttta

tctagtgttatcctctatcttcccaaccgttgggtgcacttatgatgttcattctcatt
gtctttacgattgagttgatcggtgaatttactggattaatagatgataattatcgacca
atattaaatttaattacaaaaaaaggaaatggtaagcaacatcattatgatgagcatcac
cgttcataa

5

Sequence 1910

VIIGSSSLDSSQLQALYETLYMVTVSLVIGALIGIPLGILLVVTRKNGIWSNTILHQVL
NPIINILRSIPFIILLIAIVPFIIIVISKKLDLVDRPNFRKVHTKPISVMGGTVILFSFL
IGIWLGHPIEREVKPLILGAITMYMVGLIDDIYDLRPLYKLQAGQIVAAALIVTFYGITIDF
10 ISLPIGPTIHFIFSIPTVIWIVAITNAINLIDGLDGLASGVSAALMTIGFIAILQAN
IFIIMICCVLLGSLGLFLFYNFHPAKIFLGDGALMIGFIIGFLSLLGFKNITFIALFFP
IVILAVPFIDTLFAMIRRMKKGQHIMQADKSHLHHKLLALGYTHRQTVLLIYSIAIMFSL
SSVILYLSQPLGALMMFILIVFTIELIVEFTGLIDDNYPILNLITKKGNKGQHHYDEHH
RS*

15

Sequence 1911

Contig_0678_pos_4650_3682,

is similar to (with p-value 1.0e-43)

>sp:sp|P30363|ASPG_BACLI L-ASPARAGINASE (EC 3.5.1.1) (L-ASPA
20 RAGINE AMIDOHYDROLASE). >pir:pir|S18999|S18999 asparaginase
(EC 3.5.1.1) - Bacillus licheniformis >gp:gp|Z11497|BLANSAG_
2 B.licheniformis ansA gene for asparaginase. NID: g49270.
atgaaacgtctacttatcacatactggtggcacaataagtatgtcacaagatcaaact
aataaagtgataacgaatgaagaaaatccaatatcacaacatcaaaatatcattagtcaa
25 tatgcagagggttgacgaatcaatcttttaatataccctcgccgcatatgacaatttcg
aatgttgtgcgattaaagagacgaaatcattacatatcttaagaaaatatatgatgga
tttgtcattactcatggaacagatacacttgaggaaacagcttttttaatatagatttatta
attgatattcaagagcctatagtaattactggagcaatgagatcatccaatgaaattggt
tccgatggtctctataattttatttctgctataagggttgcttctcatctgaggcta
30 cataaagggtgttatggtcgtatttaatatgatgagattcacactgctcgtaatgtgacaaag
acacatacttcgaatattaatacatcttcaaagtcctaatacaggggcctctaggtgtactt
accaagaatcgagtacaattttatcatcatccttacagacaaactacctaccaatatac
gatgtaaatttacctgtgtccacttgtaaaagcatacatgggtatggaagatgatgtacta
tcattttattcacacaacacggttgatggtatagtcacgaagcactaggacaaggtaac
35 cttccaaaaagttgtcttaattggactacagcaatgtctaaagaaaaacattcctctagtt
ctcgtatctagatcattcaatggtattgttagtcctgtatatgcttatgaagggtggtggc
gcagatttgaaaaataatggtgtatttttgcgaacggtttaaatggacaaaggcaagg
ctaaaattactagttggtttgagtcaagacatgactcaaaatcaattagagcgatatttc
gaagagtaa

40

Sequence 1912

MKRLIIHTGGTISMSQDQTNKVITNEENPISQHQNIISQYAEVDEINLLNIPSPHMTIS
NVVRLRDEIITYSKENIYDGFVITHGTDLTLEETAFLIDLLIDIQEPVITGAMRSSNEIG
SDGLYNFISAIRVASSSEANHKGMVVFNDEIHTARNVTKTHTSNINTFQSPNQGPLGVL
45 TKNRVQFYHHPYRQTTYQYIDVNLRVPLVKAYMGMEDDVLSFYSSQQHVDGIVIEALGQGN
LPKSCNLGLQQCLKKNIPVLVSRSFNGIVSPVYAYEGGGADLKNNGVIFSNGLNGPKAR
LKLLVGLSQDMTQNLERYFEE*

Sequence 1913

Contig_0681_pos_393_1478,

is similar to (with p-value 0.0e+00)

>sp:sp|Q24803|ADH2_ENTHI ALCOHOL DEHYDROGENASE 2 (EC 1.1.1.1
(ADH) / ALCETALDEHYDE DEHYDROGENASE (EC 1.2.1.10) (ACDH).
>gp:gp|U04863|EHU04863_1 Entamoeba histolytica HM1:IMSS alco
55 hol dehydrogenase 2 (EhADH2) mRNA, complete cds. NID: g48842
9.
gtgctgagacgccgagaaaaccaaccacaaatcaaagtgtttaacgaagttgaaccta
ccatcaactcatagctctataagggttagaaatgtttataaatttccaacctaatact
attattgcactcgggtggcggttcggcaatggatgcagccaaagcaatatggatgttcttt

gagcatccagaaacttcattttttggggcaaaacaaaagttcttagatattcgtaaactg
 acttataaaattaccaaacctaaaaacgaaaatttatatgtataccaacgacatcagga
 actggttctgaagtgcacacctttgcagtaattactgatagcgagacacagtttaagtat
 ccactagcagattatgctgtaactcccgatattgctatcgatccacaattcgtatta
 5 agtgtacctaagaatggtgcccgcagatacaggaatggatgttttgacacatgccattgaa
 tcttacgtctctgtcatggcttcagattatacaagaggcttaagcttacaagcaataaag
 ttaacttttgattatctaaaatcatcagttcaagaaaatgacaaacactcacgagaaaaa
 atgcataatgcttcaacaatggccggtatggcatttgccaatgcttttttaggaatttct
 cattctatcgcacataaaaattggtggtgaatatggtattccccacggcagaacaaatgct
 10 attttattaccacatgtcattcgctataatgccaaagatccacaaaaacatgcactgttt
 cctaaatatgatttcttttagagcagatactgactatgctgacattgcaaaattttttagga
 ctcaaaggaatacaactgaagaatttagtggtgctctagctaattgcggtgatgattta
 ggatgttcagttggtattgatatgaatttaaaatcacaaggcgtaactgaagagcttctt
 cactactatagacagaatggctgaatttagcatttgaagatcaatgtacaactgcta
 15 ccaaaagaaccgctaatttagtgaacttaaggcattatcgaaacagcatatgattatgaa
 agataa

Sequence 1914

VLRRRENQPQIKVFNEVEPNPSTHTVYKGLEMFINFQPNTHIALGGGSAMDAAKAIWMFF
 20 EHPETSFFGAKQKFLDIRKRTYKITKPKNAKFICPTTSGTGSEVTFPAVITDSETHVKY
 PLADYALTPDIAIVDPQFVLSVPKDVAADTGMDVLTHAIESYVSVMASDYTRGLSLQAIK
 LTFDYLLKSSVQENDKHSREKMHNASTMAGMAFANAFGLISHSIAHKIGGEYGIPHGRNTNA
 ILLPHVIRYNAKDPQKHALFPKYOFFRADTDYADIAKFLGLKGNTTEELVDALANAVYDL
 GCSVGIDMNLKSQGVTEELLHSTIDRMAELAFEDQCTTANPKEPLISELKGIIETAYDYE
 25 R*

Sequence 1915

Contig_0681_pos_4198_4911,
 is similar to (with p-value 2.0e-32)
 30 >gp:gp|AF008930|AF008930_3 Bacillus subtilis choline transpo
 rt system including ATPase (opuBA), transmembrane protein (o
 puBB), choline binding protein precursor (opuBC) and transme
 mbrane protein (opuBD) genes, complete cds; and unknown gene
 . NID: g3068551. >gp:gp|Z99121|BSUB0018_58 Bacillus subtilis
 35 complete genome (section 18 of 21): from 3399551 to 3609060
 . NID: g2635827.
 gtgtcttcttcaataagatattttatacatattcgtaataattgacggttctgaacctagc
 ttgcctgcgaatgtgattttatcacctttttgcgctgccataggtatggcaatagctatg
 ataatacacaattacaattgtccctaaagaaatgagcaattttttatatgataaacgttcc
 40 atgtatcttaaaataaaatcaaaaataatagctagaagtgcagctggaatagcacctatt
 aaaatgagtgcaactattgttacgatcaatgcctaataatattaaatctcctagaccacca
 gcgcctatttaaagctgcgagtgtagcagtagccaatgattaataccatagctgtgcgtatt
 cctgccatgataacaggcattgcaataggaggttcgactttggtcaatcttctaagtggt
 ttcattccaatgccttttagccgcttcaataagagagggatcgacctccttaataacctgtg
 45 tatgtattacgtagaataaggaagtaacgcatatacaactaaggcgataattgccggaagt
 ctcccaattccgaaaattggtatcatcaaaccaagtagtgccagtgatggaattgtttgt
 agaacagctgcaatattcatgacgatttctgaaagttttttgttttagtgagcaaaatt
 gcaattggtactgctataagagttgcaatgaaaagcgctataaaagaaagttga

Sequence 1916

VSSSISILYIFVIIDGSEPSLPANVILSPFCAAIGMAIAMIITITIVPKEMSNFLYDKRS
 MYLKIKSKIARSAAIGAPIKMSALLLRSMPPNNIKSPRPPAPIKAASVAVPMINTIAVRI
 PAMITGIAIGSSTLVNLLSGFIPMPLAASIREGSTSLIPVYVLRIGSNAYTTKAIAGS
 50 LPIPKIGIIPSSASDGIVCRTAAIFMTISESFFVLVSKIAIGTAIRVAMKSAIKES*
 55

Sequence 1917

Contig_0681_pos_5858_4974,
 is similar to (with p-value 2.0e-61)
 >gp:gp|AF008930|AF008930_2 Bacillus subtilis choline transpo

rt system including ATPase (opuBA), transmembrane protein (opuBB), choline binding protein precursor (opuBC) and transmembrane protein (opuBD) genes, complete cds; and unknown gene . NID: g3068551. >gp:gp|Z99121|BSUB0018_59 Bacillus subtilis complete genome (section 18 of 21): from 3399551 to 3609060 . NID: g2635827.

5 gtgttaattggtccttcagggttggtgaagacaacaacacttaaaatgatcaatcgttta
 attcctttaagcgaaggttacatatattttaagataaaccaattagtgattatccagtt
 tacgaaatgcgttgggatattggatatgtactacaacaaattgcattatctcctcatatg
 10 actatcaaagaaaacattgcacaagtaccccaaatgaaaaagtggaaagataacgatatt
 gatttacgtgtggatgaattattgacaatggttaggtttaaatcctgagcaatttaaaaaat
 agaaaacctgatgaactatcgggtggacaacgacaaagggttggtgtagtgcgagcatta
 gctgcagatcctcccgttattatcatggatgagccttttagtgctcttgaccaattagc
 agagagaagttgcaggacgatttgattcatttgcaaactcaaattaagaaaacaatagtt
 15 tttgtaacgcatgatattcaagaagcgatgaaactaggtgacagaatttggttactcaat
 gaagggcgtgttgaaacaaattgatacacctgatagtttagaacgcgacctaaaagtgc
 tttgtaaagcaattcatgggtagtcatttggtatcaacgcataatattgtgaatcaagt
 aagattaagatttaggaattaatcggctctgtagatgacaatgcaagcgattatcattat
 cctgaagttagtgctgagttgtacttaataatatttatgaagatttgctcattatgat
 20 gcagttgtgtcaacgataaagaacaacatcgctcgatatttgtaaatagagaagatgtg
 tttacttatttatctcttaataaggaggaagcgacacatgaatga

Sequence 1918

VLIGPSGCGKTTTLKMINRLIPLSEGYIYFKDKPISDYPVYEMRWDIGYVLQQIALFPHM
 25 TIKENIAQVPQMKKWKDNDIDLVRDELLTMVGLNPEQFKNRKPDELSGGQRQRVGVVRL
 AADPPVIIMDEPFSALDPISREKLQDDLIHLQTQIKKTIIVFVTHDIQEAMKLGDRICLLN
 EGRVEQIDTPDSFRTRPKSDFVKQFMGSHLDTTHNIVNQVKIKDLGINRSVDDNASVIHY
 PEVDAELYLNNIYEDLSHYDAVVVNDKEQHRRYLLNREDVFTYLSLNKEEATHE*

Sequence 1919

Contig_0681_pos_4819_3467,
 is similar to (with p-value 5.0e-36)
 >gp:gp|AF008930|AF008930_4 Bacillus subtilis choline transpo
 rt system including ATPase (opuBA), transmembrane protein (o
 35 puBB), choline binding protein precursor (opuBC) and transme
 mbrane protein (opuBD) genes, complete cds; and unknown gene
 . NID: g3068551. >gp:gp|Z99121|BSUB0018_57 Bacillus subtilis
 complete genome (section 18 of 21): from 3399551 to 3609060
 . NID: g2635827.

40 atgaatattgcagctgttctacaaacaattccatcactggcactacttggtttgatgata
 ccaattttcgggaattgggagacttccggcaattatcgcttagttgtatatgcgttactt
 cctattctacgtaatacacacaggtattaaggaggtcgatccctctcttattgaagcg
 gctaaaggcattggaatgaaaccacttagaagattgaccaaagtcgaactccctattgca
 atgcctgttatcatggcaggaatacgcacagctatggattaatcatttggtactgtaca
 45 ctcgcagctttaataggcgctggtggtctaggagatttaataattattaggcattgatcgt
 aacaatagtcgactcattttaataggtgctattccagctgcacttctagctattattttt
 gattttattttaagatacatggaacggtttatcatataaaaaattgctcatttctttagg
 acaattgtaattgtgattatcatagctattgccatacctatggcagcgcaaaaaggatg
 aaaatcacattcgcaggcaagctaggttcagaaccgtcaattattacgaatatgtataaa
 50 atacttattgaagaagacacagatgatactgtagaagtcaaagatggcatgggtaaaacc
 tcattcttatttaatgcgcttaagtcagatgaaattgatggtatttagaatttacaggt
 actgtattagggtgaattaacgaaagaagatttaaagtctaaaaaagaaaacgatgtatat
 caacaagcaaagtctagtttagaaaaaaaatatgatatgacaatgcttaaacgatgaaa
 tataataatacgtatgcttagctgtgtaaaacgtgactttgcaaaaaaatatcaaattaag
 55 acaataggtgatttacgcaaggtagaagataaaacttaaacaggttttacattggaattt
 aatgatagaccagatggatacaaagctgttaaaaaacgtatcatcttaattctttctaat
 gttaaaactatggaacctaaattacgttatactgcagttaaaaaggagatattaatctc
 atagacgcatactctactgatgcagaattaaacaatatatacatggtagtattaaaagat
 gatcaacatgtatttctccataccaaggagcaccgctattttaagaaaaatattttaaa

gaccatcctgaagttaaaaaacgctcaataaattggcgaatagaatcacagatgaagaa
atgcaagaaatgaactataaggttaacagtgaagaaaggagatccttataaagtagcaaga
gaatacttagaaaaaagaaaaattaataaaaaataa

5 Sequence 1920
MNIAAVLQTIPTSLALLGLMIPIFGIGRLPAIIALVVYALLPILRNTYTGIKEVDPSLIEA
AKGIGMKPLRRLTKVELPIAMPVIMAGIRTAMVLIIGTATLAALIGAGGLDLILLGIDR
NNSALILIGAI PAALLAII FDFILRYMERLSYKKLLISLGTIVIVIIIAIAPMAAQKGD
10 KITFAGKLGSEPSIITNMYKILIEEDTDDTVEVKDGMGKTSFLFNALKSDEIDGYLEFTG
TVLGELTKEDLKSKEKENDVYQAKSSLEKKYDMTMLKPMKYNNTYALAVKRDFAKKYQIK
TIGDLRKVEDKLKPGFTLEFNDRPDGYKAVKKTYHLNLSNVKTMPEPKLRYTAVKKGDINL
IDAYSTD AELKQYNMVLKDDQHVFPPYQGA PLFKEKYLKDHPEVKKPLNKLNRITDEE
MQEMNYKVTVKKEDPYKVAREYLEKEKLIK*

15 Sequence 1921
Contig_0681_pos_2086_1547,
putative peptide of unknown function
atgacgcgacagagaatcgccattgatatggatgaagtgccttgctgatacattgggtgct
gttggttaaagcgggtcaatgaacgagcggattttaaatacaaaatggaatcattaaacggt
20 aaaaaattaaaacatatgatacccgagcatgaggggttagtcatggatattttaaagaa
cctggattctttagaaatttagatgtaatgccgcatgctcaagaagttgtaaaacaactc
aatgagcattacgacatatacatagccacagcagcagatggatgttccaacctcttttcat
gacaaatatgaattgggttattagaatactttccttttttagatccgcaacattttgtattt
tgtggttagaaagaatattattcttcgagattatcttattgatgataatccaaagcaattg
25 gaaatttttgaagggaatcaattatgtttactgcttctcataatgttaatgaacataga
tttgaacgcgtaagtgggtggagagatgtaaaagaattattttaattcaattgaaaaatag

Sequence 1922
30 MTRQRIADMDVLA DT LGAVVKAVNERADLN IKMESLNGKKLKHMIPEHEGLVMDILKE
PGFFRNLDVMPHAQEVVKQLNEHYDIYIATAAMDVPTS FHDKEYEWLLEYFPFLDPQHFVF
CGRKNII LADYLIDNPKQLEIFEGKS IMFTASHNVNEHRFERVSGWRDVKNYFNSIEK*

35 Sequence 1923
Contig_0683_pos_7975_7574,
putative peptide of unknown function
atgataaaaaaatagaacacaatcgcaaaaaacaacaaaatgatacttcaaatacaaat
cgtgataactaatcaacaccaagaccacaaactcaaccaacaaataatgactataacaacgat
40 aatcaatcaggtactgaacaaccagcacaacaacctaactaccatcaatacccaataat
aatcaacagctcgtgttcaaataaaaaataactcttcagaaaaatacaaacagaaaccgaat
cagaacaaaactaatcaatcatatcatcaaccagcacaatcaacaccacaacagtcgtca
caacataataatcaatctgattcacaacaaaatggcaactcaaataataattccaacaat
caaaatcatggaacaaatgataaacagaataaaaaatcgtaa

45 Sequence 1924
MIKKIEHNRKNKQNDTSNQNRDTNQHQDQTQPTNNDYNNDNQSGTEQPAQQPNYHQYPNN
NQQSGSNKNNSENKQKPNQKTNQSYHQPAQSTPQQSSQHNNQSDSQONGNSNNNSNN
QNHGTNDKQKNR*

50 Sequence 1925
Contig_0683_pos_6967_6455,
putative peptide of unknown function
atgggttggaacgggtgttaagtgggttttgaatataaagcacaaaaagaaaaatagataac
55 ttaacaaagtctctttaaagacaatgaagaaaaatatcaatataccggttttactaaagaa
gcaatacataagacacagaatgttgatatcaaaatgagtattattatttagcaggtaac
gttactaatattaataattatagaaaatattatgaacctttaataaaaaaagattctaag
aatttcaaagaaggcatgaaaaagcaaatgaagcaacaaatttcaaagccaaaattgaa
gttgtttcaacattatttcagtactaaatctgatttcactaaaaataactctaagaaagat

ttattattcttaagtgatgatttatatcattacaaagaaaaacctgaaaacacaaacata
actttacaattaagttagccaaaaattaattctacacgcgcattttatgatgctaacaac
ccattagaatatggagtgcataaacatgagtaa

5 Sequence 1926

MVGTVLSGFEYKAQKEYDNLTKEFKDNEEKYQYTGFTEAIHKTONVGYQNEYYLGN
VTNINNYRKYEPLIKKDSKNFKEGMKKANEATNFKAKIEVVSTLFSTKSDFTKNNSKKD
LLFLSDDLYHYKEKPENTNITLQLSEPKINSTRAFYDANNPLEYGVHKHE*

10 Sequence 1927

Contig_0683_pos_5976_5449,

putative peptide of unknown function

atgatggtaggaactgtgttaagcgggttttgagtatagagcaaataaggaaaaaatggat
aacttagaaaaatacctcaaagataaagaagataaatatcactatactggattcacccgat
15 gaagcaataactaaaactcaaaatataggttatcaaaataattacttttacattactact
agttccacgaaattacgagattatagaaaacattttgaacctttaattaaagaaagtgc
gatgattttaaaaagcatatgaaacaattaaagtctaaaaaagatacgtatattaataca
gaaataacgactacacttttcagtactctggacgaatatgacgaaaaaatcattagaaaa
ataactttatccatggctaaagaaatgagaaaaagagccatctattccacataatttcaca
20 ttocactttatttttagcaataacaaattaaaaatcaacgatccaaacataagtaacaat
caaattaatgagtataggggtgttcgaccatgacggatttaaaaaattaa

Sequence 1928

MMVGTVLSGFEYRANKEKMDNLEKYLKDKEDKYHYTGFTDEAITKTQNIQYQNNFYITT
25 SSTKLDRYRKHFELIKESDDDFKKHMKQLKSKKDTYINTEITTLFSTLDEYDEKIRK
NTLSMAKEMRKEPSIPHNFTHLLFSNNKLKINDPNISNNQINEYRVFDHDFGN*

Sequence 1929

Contig_0683_pos_3077_1491,

30 is similar to (with p-value 2.0e-17)

>sp:sp|P46321|CELR_BACSU PUTATIVE CEL OPERON REGULATOR. >pir
:pir|S57758|S57758 probable cel operon regulator - Bacillus
subtilis >gp:gp|Z49992|BSCCLABCD_1 B.subtilis celA, celB, ce
lC, celD and ywaA genes. NID: g895746. >gp:gp|Z99123|BSUB002
35 0_155 Bacillus subtilis complete genome (section 20 of 21):
from 3798401 to 4010550. NID: g2636240. >gp:gp|D83026|D83026
_60 Bacillus subtilis genome sequence covering lic-cel regio
n. NID: g1783231.

atggaaaagcatattcactatcaaagagctatcttcaacttaccatttaacaaaatca
40 aagtgtattgattatgaacacgtatatacaaacgtgggtataaaatttgatattttta
tcaattaaaaagaagcaaggtatcatgatcgatgcgagtagacaacgagtagatcagtaatgct
gtacttcatatcaatcaacttacagacgatgactttaaagttgaaaaccttattttacaa
gagttacctcaagccatactagaaaaataaaacaaattatctcaaagcatatagataat
catcaattatcaacttctgaaaataaaatatacaaatcttctgcatctaattttaatt
45 atcaaacattctcaaccagaggaagaagattggagcactgatacagaatctttaactatt
gcgaaaaagtgataaaagatatcaatgaaacccttgatatcaacttaacaataaaaca
agtgaatgcttttcttttttatttagctaccatttcaataagtttgatttagggatccaa
caactattttattcaagtttatatcgatcgactcattgaattaatggagcaacatatgggt
tttcccttttcacaagatacaatttttaaagataatatgaacgtccacttttagtcgtaca
50 tatttgcgattaatgagtcattgtttatctaaataatccattaacaagtcaaatcaaacga
ctatatccctttgtcttttaatacactatatgatagattcgacaattatcacaagatacc
aatatccaattaagcgaagatgagattgcctttttaactatacattttcagtccttctatc
gaacgccataagtcatcacatattcatgttgtaattgcttggtattatggcttaggcatt
tcaacgttgcttgctgagaaaatcaacaacttaacatcatgaatacagatcgtagataca
55 taaaacttgaagattataaataactatcattttgaaggattgacttattaattactact
cacgactttgatatacagtcacttttacaaatacctaaagtcatacaagtatcaccttta
ttttcagatgaagatgctaaaaaatcgaattctttgtaaaagccatgcaaaaccatta
tcaaaagatgatataattatcaaaaattcagttgagtggtgagtcgaatttcaactgaat
cattcaaatcacattcttccaatttttgagaaatccaaagaaatttagattatcatcat

gcaactctagatggctatatagaaaagtgccatagatcgcgaaaaacaatcttcaacatat
 ataggtaaagggatagcactcccacacggcaaccctgaaaaagtactgaaatcacacatg
 attatatttaaaccttctcaacctataacatggaacaacatgaagttaaacttggtttc
 ttttagcaatgagtaaaaaagatttaaatattaaccgtaaaattatacaatcaattgct
 5 caattagaagaagatgacatccatcaattatgtcttttagatgattacaactaaaaaac
 actttgtatgcacgttttaagaataa

Sequence 1930

MEKHTFTIKELSSYHLTKSKVIDYVTRIQTWAIKFDIYLSIKKKQGIMIDASTTSISNA
 10 VLHINQLTDDDFKVENLILQELPQAHTRKIKQIISKHIDNHQLSTSENKIQQLLVHLILI
 IKHSQPEEEDWSTDESILTIAKKCIKDINETLGYQLNNKTSECFSSFFISYHFNKFDLGIQ
 QLFIQSYIDRLIELMEQHIGFPFSQDTILKDNMNVHFSRTYLRLMSHVYLNPLTSQIKR
 LYPFVFNLTLYDSIRQLSQDTNIQLSEDEIAFLTIHFQSSIERHKSSHVVIACYYGLGI
 STLLAEKIKQLNHAIQIVDTLKLEDINNYHFEGIDLITTHDFDTSQLLQIPKVIQVSL
 15 FSDDEAKKIEFFVKAMQNPLSKDDILSKIQLSVESNEFKLNHNSHILPIFEKSKEILDYHH
 ATLDGYIESAIDREKQSSTYIGKGIAPHGPNPEKVLKSHMIIFKPSQPITWKQHEVKLVF
 FLAMSKKDLNINRKIIQSIAQLEEDDIHQCLLDDLQKNTLYARFKE*

Sequence 1931

20 Contig_0684_pos_1289_1969,
 is similar to (with p-value 3.0e-25)
 >sp:sp|Q49435|Y442_MYCGE HYPOTHETICAL PROTEIN MG442. >pir:pi
 r|H64248|H64248 hypothetical protein homolog MG442 - Mycopla
 sma genitalium (SGC3) >gp:gp|U39731|U39731_1 Mycoplasma geni
 25 talium BS17, pilB_2, rpl19, trmD genes from bases 546767 to
 554372 (section 53 of 56) of the complete genome. NID: gl046
 159. >gp:gp|U39726|U39726_4 Mycoplasma genitalium section 48
 of 51 of the complete genome. NID: g3845031.
 atgacgaatttaaaagaattagaaaaatgggaaacttatttttaaagatgaaggtttctat
 30 ccggtagctgtagatgcaaaacatggcaagaatcttaaaaatgttgaagttgaagctata
 aaagcaactcaagaaaaatttgatcgtgaaaaagctaaaggtttaaacctagagcgata
 agagctatgattgttaggcatttcctaatgtaggaaaaatcaacacttatcaataagtttagca
 aaacgtagtatcgccgaaactggaaataaaccaggagtaacaaaacagcaacaatggatt
 aaagtttgaaagtctcttcaattactagatacaccagggtattttatggcctaaattcgaa
 35 gatgaagaagtcggtaaaaaattaagtttaactggtgcaattaaggatagtagtgcgttc
 ttagatgaggtagctattttatggtttgaattttatgattaaacatgatgtttcagcttta
 aagagacattataatattgatacacatgaagacgctgagatactcgattgggttgatgca
 attggaagaagaaggggattgttacaaaaaggaaatgaagtagattatgaatctgtcatt
 gagttgatcattcaatgatagataatgcaaaaattggaacttattgttttgatatttta
 40 aaagaaatgaagagtgaatga

Sequence 1932

MTNLKELEKWETFYFKDEGFYPVAVDAKHGKNLKNVEVEAIKATQEKFDREKAKGLKPRAI
 RAMIVGIPNVGKSTLINKLAKRSIAETGNKPGVTQQQWIKVGKSLQLLDTPGILWPKFE
 45 DEEVGKKLSLTGAIKDSIVHLDEVAIYGLNFMIKHDVSAKRRHYNIOTHEDAEILDWFDA
 IGRRLGLLQKGNVDYESVIELIINDMRNAKIGTYCFDILKEMKSE*

Sequence 1933

Contig_0684_pos_1974_2744,
 50 is similar to (with p-value 9.0e-41)
 >gp:gp|AF005098|AF005098_1 Lactococcus lactis RNaseH II (rnh
 B) gene, partial cds, positive regulator GadR (gadR), GadC (
 gadC) and glutamate decarboxylase (gadB) genes, complete cds
 . NID: g2352483.
 55 atgtctctaacaattaaagaaatcaaagaaaaactatctcgaattgaaacggttggaagag
 ttacataaacatgaagcaataatgattcacgtaaggtgttataaatgcgattaagtct
 agggaaaaaaatattcttaagcaacaagcattagaagagcactatttatccatgaatcaa
 tacgaaaaaacattatgtcctctaacagggatgcattaatttgtggaattgatgaggta
 gggcgtgggccccttggtggtgaccagttgtggcttggtgcagttattttagagaagaatcat

cattatattgggttagatgactctaaaaaagtgtctccaaaaatagagcagcacttaat
caaaatttaaaagaaaatgtctatcaatatgcatatggcatagcgtcctcagttgaaata
gatgaattgaacatttatcgggcaactcaattagctatgctacgtgctataaatcaatta
gatgttacacctacacattttatataagacgcaatgacactagatattgatattccacaa
5 acctcaattattaaaggtgatgctaaaaagtgtgtctatcgagcagcaagtatcatggct
aaagaataccgtgatcaatatatgagacaactatctaaacagtttccagaatatggtttt
gataaaaaatgcaggttatggaactaagcaacatttaaaggctattgatcaagtgggcata
atcaatgaacatcgatcaatcatttgaaccaattaaatcaatgatgaaataa

10 Sequence 1934

MSLTIKEIKEKLSRIETLEELHKHEANNDSRKGVINAISREKNILKQQALEEHYLSMNQ
YENNIMSSNRDALICGIDEVGRGFLAGPVVACAVILEKNHHYIGLDDSKVSPKNRRLN
QNLKENVYQYAYGIASSVEIDELNIYRATQLAMLRAINQLDVTPTLLIDAMTLDIDIPQ
TSIIKGDAKSVSIAAASIMAKEYRDQYMRQLSKQFPEYGFKNAGYGTKQHLKAIDQVGI
15 INEHRQSFEPKSMMK*

Sequence 1935

Contig_0684_pos_2945_4018,

is similar to (with p-value 0.0e+00)

20 >gp:gp|Z99112|BSUB0009_79 Bacillus subtilis complete genome
(section 9 of 21): from 1598421 to 1807200. NID: g2633902. >
gp:gp|AJ000975|BSYLQGCOD_3 Bacillus subtilis ylgg to codV ge
ne region. NID: g2462964.

25 gtggaaaaagcgaaagaattaaattcagacgtatatgtgggttaaagcgcaaattcacgct
gggggttagaggttaaagcagcgcgctgaaaaattgctaaatcattatctgaagtcgaaacg
tacgcaaatgaactgctaggtaaacaattggtcacacatcaaaactgggccagagggcaaa
gaggtcaaacgtttatatatcgaagaaggatgcatatccaaaaagaatattatgttggt
tttgttattgatcgctgctactgataaagtactttgatggcatcagaagaaggtggaact
gaaattgaagaggttgacgctcaaacacctgaaaagattttcaaagaaacaattgatcca
30 gtagtaggattatcaccttaccaagcgcgacgtatcgcttttaataattaacattccaaaa
gaatcagttggaaaaagcaactaaatttttattagcactatataatgtctttatcgaaaaa
gattgttctattgttgaaattaaaccacttggtacaactggagacgggtcaggtattggct
ttagatgctaaattaaactttgatgataatgcattatttagacataaagatatttagaa
ttacgagatttagaagaagaagatcctaaggaaatagaagcttctaaatatgatttatca
35 tacatcgctttagatggagatattgggttgatgggttaatggcgaggtttagccatggca
actatggatacaattaatcattttgggtggaatccagccaacttcttagatgtaggtggc
ggtgctacaaaagaaaaggttaactgaagcatttaaaattatttttaggtgatgacaatgtt
aaaggtatccttgttaaatatttttggtggaattatgaaatgtgatgttattgccgaaggt
attgtgacagcggtttaaagaagttgaactaacattaccattagttgttcgtttagaagga
40 actaatgtcgacgtggtaaaagcaatattaaacgaatcaggttttagctattgagccagca
gcaactatggctgaaggtgctcaaaaaattgtgaaacttggttaaagaagcataa

Sequence 1936

45 VEKAKELNSDVYVVKAQIHAGGRGKAGGVKIAKSLSEVETYANELLGKQLVTHQTGPEGK
EVKRLYIEEGCDIQKEYYVGFVIDRATDKVTLMASEEGGTEIEEVAAQTPEKIFKETIDP
VVGLSPYQARRIAFNINIPKESVGKATKFLALYNVFIKDCSIVEINPLVTTGDGQVLA
LDAKLNFDNALFRHKDILELRDLEEDPKEIEASKYDLSYIALDGDIGCMVNGAGLAMA
TMDTINHFGGNPANFLDVGGGATKEKVTEAFKIILGDDNVKGIFVNIFGGIMKCDVIAEG
IVA AVKEVELTLPVVRLEGTNVERGKAILNESGLAIEPAATMAEGAQKIVKLVKEA*

50

Sequence 1937

Contig_0684_pos_4121_4645,

is similar to (with p-value 1.0e-69)

55 >gp:gp|Z99112|BSUB0009_80 Bacillus subtilis complete genome
(section 9 of 21): from 1598421 to 1807200. NID: g2633902. >
gp:gp|AJ000975|BSYLQGCOD_4 Bacillus subtilis ylgg to codV ge
ne region. NID: g2462964.

atgcttgattatgggactcaaattgttgacggggtaaacacctggtaaaggtggacaagtt
gtagaaggtgttccagtatataacactgttgaaagaagctaaaaatgaaacaggagcta

gtatctgttgatatacgtaccagcaccattcgctgctgattcaattattgaagcagctgat
 gccgatttagacatgggtatttgtattactgaacatatacctgttgatgatgggttaa
 gtaaaaagatatattacaaggtcgtaaaacacgttttagtaggaccaaactgtcctgggtg
 ataactgccgacgagtgtaaaatcggtattatgccgggatatatccataaaaaaggccat
 gtcggtgtcgtgtctcgttctggtacattaacgtatgaggcagtgcatcaattaactgaa
 5 gaaggatcggtcaaaactgctgtaggtatcggcggtgatccagtaaatgggactaac
 tttatttgtctaatcatagacaaaagtgtcatctctcattctag

Sequence 1938

10 MLDYGTQIVAGVTPGKGGQVVEGVPVYNTVEEAKNETGANVSVVYVPAPFAADSIIEAAD
 ADLDMVICITEHIPVVDMMVKRYLQGRKTRLVGPNCPGVITADECKIGIMPGYIHKKGH
 VGVVSRSGTLTYEAVHQLTEEGIGQTTAVGIGGDPVNGTNFICLIIDKSAHLSF*

Sequence 1939

15 Contig_0684_pos_8692_9123,
 is similar to (with p-value 2.0e-20)
 >gp:gp|U96107|SCU96107_3 Staphylococcus carnosus N5,N10-meth
 ylenetetrahydromethanopterin reductase homolog, SceB precurs
 or (sceB) and putative transmembrane protein genes, complete
 20 cds, and putative Na⁺/H⁺ antiporter NhaC (nhaC) gene, parti
 al cds. NID: g2735503.
 atgaaattcaaaaaattattatctcgtattattatcgctacaatgattacatttactgga
 acactctcatatcaagctattgaacaaacgcataatttcccatgctgcacataattattat
 ggtaaaaaacaatgcacttgggtgggcattttaaactcgtgctcaattaggtaaactgta
 25 tcaaatcgatgggtaattgctaagaattggtatagcaatgcacgtcgatctggttatgca
 atggacataagcctcgaaaatacgtctgttatgcaatcaacgagaggctattatgggcac
 gtagcagtggttgaaaaagtataagaatggaaaaatcaaaatttctgaatataattat
 aatgtgccattaggctacggcacacgcattattagtaaatcgtctgcacgaaactataat
 tatattttat

Sequence 1940

30 MKFKKLLSRIIIATMITFTGTLISYQAI EQTHISHAAHNYGKKQCTWWAFKRRRAQLGKPV
 SNRWGNNAKNWYSNARRSGYATGHKPRKYAVMQSTRGYYGHVAVVEKVYKNGKIKISEYNY
 NVPLGYGTRIISKSSARNYNYIY*

Sequence 1941

35 Contig_0684_pos_9983_0,
 is similar to (with p-value 1.0e-85).
 >gp:gp|Z99122|BSUB0019_80 Bacillus subtilis complete genome
 40 (section 19 of 21): from 3597091 to 3809700. NID: g2636029.
 >gp:gp|Z92954|BSZ92954_7 B.subtilis yws[A,B,C,D,E,F,G] and g
 erBC genes. NID: g1894764.
 atgaataaattaggctcgtcgacgtctagtaatgctcatcgctatagttattatttggg
 gcaactgattcttgcgtcatcaaccaatttagcattattaattattggacgtttaattatt
 45 ggttttagcagtggttggttcgatgtctacggtacctgtttatttaagtgaatggcaccg
 acagaatatcgtggctcactagggtcacttaataactaatgattacaattggattttta
 gcagcatatttagtcaactatgcatttgcggatattgaggggtggcgttggtatgctagga
 ttagcagttgttccatcggtattttacttgtgggtattttatttatgcctgagagtcca
 agatggttacttgaaaatagaacgaagaagccgctcgtcaagtgatgaagattacttat
 50 gacgatagcgaaattgataaagaacttaagagatgaaagaaattaacgctatctctgaa
 tctacatggacagtcattaaatcaccatggttaggtagaatattaattgtaggttgata
 tttgctattttccagcaatttatttggtatcaatgcagtcattttctattcatcttcaatc
 tttgctaaggctggatgggtgaagcggcgtctatattagggttcagttggtataggaact
 attaatgttcttgaacaatagttgccattttttagtagataagattgatcgtaaaaaa
 55 ttacttgttggtggtaatttggtatgattgcctcattattaattatggcaatcttaatt
 tggacaattggaattgcttcacagcgtggattattattgtttgtttatcattatttatt
 gtattctttgggatttcttggggacctgttctatgggttatgctacctgaattattccca
 atgcgcgcacgtggcgctgctacgggcatttcagcgcttgtgctaaaatcggaaacgctt
 atcgtgtcattgttcttcccaatatta

Sequence 1942

MNKLGRRLVMLIAIVFIIGALILAASTNLALLIIGRLIIGLAVGGSMSTVPVYLSEMAP
TEYRGSLSLNLQMITIGILAAYLVNYAFADIEGWRWMLGLAVVPSVILLVGIYFMPESP
5 RWLLENRNEEAARQVMKITYDDSEIDKELKEMKEINAISESTWTVIKSPWLGRILIVGCI
FAIFQQFIGINAVIFYSSSIFAKAGLGEAASILGSVGIGTINVLVTIVAIFVVDKIDRKK
LLVGGNIGMIASLLIMAILIWTIGIASSAWIIIVCLSLFIVFFGISWGPVLWVMLPELFP
MRARGAATGISALVLNIGTLIVSLFFPIL

10 Sequence 1943

Contig_0684_pos_10911_10567,

is similar to (with p-value 3.0e-26)

>sp:sp|P46333|YXBC_BACSU HYPOTHETICAL METABOLITE TRANSPORT P
ROTEIN IN HTPG-IOLR INTERGENIC REGION.

15 atgcccgtagcagcgccacgtgcgcgcattgggaataattcaggtagcataacccataga
acaggtccccaagaaatcccaagaataacaataaataatgataaacaacaataataatc
cacgctgatgaagcaattccaattgtccaaattaagattgccataattaataatgaggca
atcataccaatattaccaccaacaagtaattttttacgatcaatcttactactacaaa
atggcaactattgttacaagaacattaatagttcctataccaactgaacctaataatagac
20 gccgcttcacccagtcacgccttagcaagattgaagatgaatag

Sequence 1944

MPVAAPRARIGNNSGSITHRTGPQEIPKNTINNDKQTIIIIHADEAIPVQIKIAIINNEA
IIPILPPTSNFLRSILSTTKMATIVTRTLIVPIPTPNIDAASPSPALAKIEDE*

25

Sequence 1945

Contig_0684_pos_9748_9200,

is similar to (with p-value 5.0e-44)

>sp:sp|P13702|MVA_PSEMV 3-HYDROXY-3-METHYLGLUTARYL-COENZYME
30 A REDUCTASE (EC 1.1.1.88) (HMG-COA REDUCTASE). >pir:pir|A44
756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88)
- Pseudomonas sp. >gp:gp|M24015|PSEHMGCOA_1 P.mevalonii HMG
-CoA reductase (mvaA) gene, complete cds. NID: g151258.

atggaacgagcgctcagttccttcacaaagtagatatacatcgtgctgcaacacataacaaa
35 ggtgtgatgaatggtatacacgctgtagatttggtacaggcaatgatacaagaggagtt
gaagcaagtgctcatgcatatgcaagcaaagatggcattatagagggatagctacttgg
gaatatgatcgtcacgtaataaattggttggaactattgaagttcctatgacttttagcg
acagtaggtggaggtagcgaagttttacatttgctaaagcctcattaaatttgcttaatt
gttgaaaatgcacaggaactagggaagttgttgctgctgttgattagcacaataatttc
40 tctgcatgtagacgctagtgctgagggatacaacaaggacatagagtttacaatat
aatcattagcgattgttgtaggtgcaaaaggcgaagaaattgcgcaagtagctgaagcg
ctcaaatatgaatcacaagctaataactgccaaagctcaagaaatcttgatgaatataaga
aagtcataa

45 Sequence 1946

MERASVLAQVDIHRATHNKGVMNGIHAVVLATGNDTRGVEASAHAYASKDGHYRCIATW
EYDRSRNKLVTIEVPMTLATVGGGTKVLPIAKASLNLNVENAQELGQVVAAGVLAQNF
SACRALVSEGIQQGHMSLQYKSLAIVVGAKGEEIAQVAEALKYESQANTAKAQEILMNIR
KS*

50

Sequence 1947

Contig_0684_pos_7905_7147,

is similar to (with p-value 2.0e-44)

>sp:sp|P39592|YWB1_BACSU HYPOTHETICAL TRANSCRIPTIONAL REGULA
55 TOR IN EPR-GALK INTERGENIC REGION. >pir:pir|S39679|S39679 hy
pothetical protein - Bacillus subtilis >gp:gp|X73124|BSGENR_
25 B.subtilis genomic region (325 to 333). NID: g413923. >gp
:gp|Z99123|BSUB0020_126 Bacillus subtilis complete genome (s
ection 20 of 21): from 3798401 to 4010550. NID: g2636240.

atggctgtccctttatttgaccggagtaaaagaagtttagtacttactgatgcaggtaaa
 atttttttcaagaaatgtcaagaaatcatcgactatatgataatttgccactgaaatt
 aatagtttgatggtttagaacagggtcatatcactattagtagtctgcagtgatgagc
 atgcgtaaatatttggcgtattaggagactttcatcaactttatccgaatattacgtac
 5 aacttaatcgaaagtgggtgtaagacgactgaaaaccttatacttaatgatgaagtggat
 attgggtgtgacaacattgccagtagatcatcaaaaatttgaatgtatatctttaacaaa
 gaagaactgactgtagttttaataaagaacatccttttagcacaaaaatcttctattaaa
 atggaagaattagctgatgagaacttcattttatttaataagatttctatctcaacgat
 aaaattattgaaaatgcgaagaatgctggattcgtgccgaacatggcctcacaatctca
 10 caatggaatgtgattgaaaatctgtcattaatcaattaggtatttccatattgccagcc
 actatagcacaattacttaatgatgacgtcaaaattgtacatttggaatatgcacataca
 acttgggagcttgggtgctcgtttggaaaaagataaaacggttaagtcatgctacaaataaa
 tggatagaatttttgaaagaagattatccgaagaataa

15 Sequence 1948

MAVPLFDRSKRSLVLTADAGKIFFKKCQEIIALYDNLPTINSLYGLETGHITISMSAVMS
 MRKFIGVLGDFHQLYPNITYNLIESGGKTENLILNDEVDIGVTLFVDHQKFECISLNK
 EELTVVLNKEHPLAQKSSIKMEELADENFILFNEDFYLNDKI IENAKNAGFVPMASQIS
 QWNVIENLVINQLGISILPATIAQLLNDDVKIVHLENAHTTWELGVVWKKDKRLSHATNK
 20 WIEFLKERLSEE*

Sequence 1949

Contig_0684_pos_6784_6443,

putative peptide of unknown function

25 atgcttattacttttataggcacagaagttcaaaaattacttcatatacctctagcaggt
 agtatcgtagggttatgcttttttctctattgttacaatttaaaattgtacctgaatca
 tggattaatgtaggagcagactttttacttaaaacaatggttttcttctttatcccatca
 gtggttaggaattatggatgttgcacataatcacgatgaattatatattattctttatt
 gttattataattgggtacatgcctttagcactatcatcaggttatatcgctgaaaaaatg
 30 ctagaaaaaagcaatacacgtaaaggaactgatcactcatga

Sequence 1950

MLITFIGTEVQKLLHIPLAGSIVGLMLFFLLQFKIVPESWINVGADFLKTMVFFFIIPS
 VVGIMDVASNITMNYILFFIVIIIGTCLVALSSGYIAEKMLEKSNTRKGT DHS*

35

Sequence 1951

Contig_0684_pos_6275_5757,

putative peptide of unknown function

40 atgaaaggtgggtacctggattaaccatgttttaaacgctacagttgtatgtcttgcatac
 ccactttatcaaaaataaaaaaagaaaataaaaaaatatttaacaatttttcacaagcgtg
 ttgactgggtgtagttctcaattttgtgttagtatttacaacgttgaaaatctttggttat
 tctaaagacacaattgttaccctgttacctagatcaattacagcagcagtaggtatagag
 gtttctcaagaattgggagggaacagatacaattactgtgctctttatcataactacaggt
 ttaatcggcagtagtttttaggttcaatgcttttacgtatgggaggttttaaatcttccatt
 45 gcgcgaggactaacttatgggaatgcttctcacgcatttgggtaccgcaaaagcattagag
 cttgatattgaatcaggagcgttcagttcaattggtagattttaacagcagtcatttagt
 tctgttctcataccagtactgattttattgttttactaa

Sequence 1952

50 MKGGTWINHVLNATVVCLAYPLYQNKKKIKKYLTIIFTSVLTGVVLNFEVLVFTTLKIFGY
 SKDITVTLPRSITA AVGIEVSQELGGDTITVLFIIITGLIGSILGSMLLRMGGFKSSI
 ARGLTYGNASHAFGTAKALELDIESGAFSSIGMILTA VISSVLIPVLILLFY*

Sequence 1953

55 Contig_0684_pos_5707_5195,

is similar to (with p-value 2.0e-21)

>sp:sp|P42405|YCKG_BACSU HYPOTHETICAL 19.0 KD PROTEIN IN TLP
 C-SRFAA INTERGENIC REGION (ORF10). >gp:gp|D30762|BACYCK_10 B
 acillus subtilis DNA around 28 degrees region of chromosome

containing yckA-H genes. NID: g710627. >gp:gp|D50453|D50453_49 Bacillus subtilis DNA for 25-36 degree region containing the amyE-srfA region, complete cds. NID: g1805369.

5 atggcacaattaaagcaaatgaagcattagtttaaggcattacaaacctggaatattgat
catttataatgggtattcctggcgactcagtagatgctgttgttgatagcttacgtacggtc
agagatcaatttaaattctctgaagatgcttcaattaaagcagcagttgaagaagcgcat
aaacatggaaaagcattgcttgttgatgatagcagtgcaaaacttagaacaacgtgct
aaagaactagatgagatgggtgcagactatatcgagttcatacaggttacgacttaca
gctgaaggaaaatctccattagacagcttgcgtacagttaaatctgttatcaaaaactct
10 aagggttcagtagcaggtggtattaaaccagatactatcaaagatattgttgctgaagat
ccagatttagttattgttggcggtattgcgaatgctgacgatcctgtagaagcagca
aaacaatgtagagcagctattgaaggtaaataa

Sequence 1954

15 MAQIKANEALVKALQTNIDHLYGIPGDSVDAVVDLSLRTVRDQFKFSEDASIKAAVEEAH
KHGKALLVDMIAVQNLEQRAKELDEMGADYIAVHTGYDLQAEGKSPLDSLRTVKSIVKNS
KVAVAGGIKPDITIKDIVAEDPDLVIVGGGIANADDPVEAAKQCRAAIEGK*

Sequence 1955

20 Contig_0687_pos_433_1173,
is similar to (with p-value 3.0e-23)
>sp:sp|P30267|YKAA_BACFI HYPOTHETICAL 50.9 KD PROTEIN IN KAT
A 3'REGION (ORF A). >pir:pir|S27491|S27491 hypothetical prot
ein A - Bacillus firmus >gp:gp|L02548|BACKATA2_1 B.firmus OR
F A and ORF B, complete cds. NID: g143118.
25 gtgccggttaacgattaataaactcttcaattttattagatcattttgtcacatggata
agtagcgcttacctcttttaactaagatattcataatgattatcattatactaggtgct
atztatccattttattaaaggacatggaatcggaataccgttgaaacaatttttagttta
tttaaagtgttgaggatcattataggcgtttgttaatttttaacattgggccaagttgg
30 ttacttaatgaacaaacgggaatgtatgttttaactatttggttaattccggtaggatta
acagtacctgcaggaggcgcggtattagctttattagtaggatatggcttattagaattt
gtaggtgtttatgcgcaaaaaattatgtaccgatatggaaaacgcctggacgttcagca
gttaatgcttttagcatctttgttgctagttttgctgtgggttacttataacgaataaa
gagtataaagaaggtaaattcacggaaaaacaagctgttatcatagcaaccggcttttct
35 acagttactgtagctttttatgatagttattgctaaaaccttaccttaatggatatatgg
aatttatatttttggtctacctgtttgttactgctgcagtaacagcttgtacagttagg
atttggcctatcagtaaaattagcaacacatatattgatcagccatttatagaagaagat
acaagcgaattaaaaggttaa

Sequence 1956

40 VPVTINNNSILLDHFVTWISSALPLLTKIFIMIIILGAIYPFIKGTWNRNTVETIFSL
FKVLGVIGVLLIFNIGPSWLLNEQTGMVFNLYLVIPVGLTVPAGGAVLALLVGYGLLEF
VGVAQKIMYPIWKTPGRSAVNALASFVAVGLLITNKEYKEGKFTKQAVIIATGFS
TVTVAFMIVIAKTLHLMIDIWNLFWSTLFVTAAVTACTVRIWPISKISNTYYDQPFIEED
45 TSELKG*

Sequence 1957

Contig_0687_pos_1231_1665,
is similar to (with p-value 5.0e-21)
50 >sp:sp|P30267|YKAA_BACFI HYPOTHETICAL 50.9 KD PROTEIN IN KAT
A 3'REGION (ORF A). >pir:pir|S27491|S27491 hypothetical prot
ein A - Bacillus firmus >gp:gp|L02548|BACKATA2_1 B.firmus OR
F A and ORF B, complete cds. NID: g143118.
gtgatgaaaaatatttgggtgaatttcaaagaaagtctgattatgactatgaatatctta
55 cccaccatattatcaatagggttaatttgccttgttactcgagaatatacagtgattttc
gattatttagcatatgtttttatccattaacttggatacttcaaataccagattccttt
ttaactgcaaaaggcgagctattggtataacagaaatgtttttgccttcatttaattgta
gtcgaagcaccattaatcactaaatttataattgctgttacttctgtttctacaattata
ttcttttcagctagtgtgcctagtattctctctactgatatacccatccgcataagagat

ttagtggttatatggtttgagagaactgtattgagtttaattatagtaacacctatcgca
tatattttttataa

Sequence 1958

5 VMKNIWLNFKESLIMTMNILPTILSIGLICLLLAETVIFDYLAYVFYPLTWILQIPDSF
LTAKGAAIGITEMFLPSLIVVEAPLITKFIIAVTSVSTIIFFSASVPSILSTDIPIRIRD
LVVIWFERTVLSLIIVTPIAYIFL*

Sequence 1959

10 Contig_0687_pos_2701_3702,
putative peptide of unknown function
atgtccttatttaaagattttttcattgcttattctaatcatatcatatcttaataaaaata
gctaataaagatgggagcctcaaatgggagcaaatcgtgtagtcgcaggaaatcagattcat
caattaattgagactatacaaatatttaaatgattataatatttcagttactgtcgactca
15 ttgggtgaatttgttaatactagagaagaaagcattaaagctaaagaagagattttagaa
attatcgatgcaatatatagcaataatgttaaggcacatatgtcagtcagataagtcaa
cttggaagtgtgtttgatttaaatcttgcttatgaaaacatgagagaaattttacttaaa
gctgataagaatgggaagatgcataatattgatacagagaagtcagatagtccttct
aaaattcaacatataattgatagattgaaaggtgaatttaaaaatgtgggtacagtcggt
20 caagcttatctgtatgaagccgatgatataattgataaatatcctgaattacgtttgaga
cttgtagaaggtgcttataaggaagatgcgtcaatcgcttttcaatcaaaagaagaatt
gacgcaaattatattagaattattaaaaaacgactactaaattcaagaactttacatcg
gtggctacacatgacaatgaaataatcaaccaagtcaaacaatttatgaaggaaaatcat
gtcagcaaagataaaatggaatttcaaattgtgtacggtttccgcacggaattagcacia
25 aaaatagctaattgaaggttattttttacagtttatgtaccatacggtaattgattggtt
gcgtactttatgagaagactagcagaacggcctcaaaactgtcattagctataaaagaa
tttactaaaccctaaatcttataaaaaggtaaccttggtataggtatatttgcaacttta
ttgacgtctcttattcttgccattaaaagacataaaaaataa

30 Sequence 1960

MSLFKDFFIALSNHTYLNKIAKKMGPQMGANRVVAGNTIHQLIETIQYLNNDYNISVTVDS
LGEFVNTREESIKAKEEILEIIDAIYSNNVKAHMSVKISQLGSEFDLNLAYENMREILLK
ADKNGKMHNIDTEKYDSLKIQHIIDRLKGEFKNVGTVVQAYLYEADDIIDKYPELRRLR
LVKGAYKEDASIAFQSKEEIDANYIRI IKRLLNSKNFTSVATHDNEIINQVKQFMKENH
35 VSKDKMEFQMLYGFRTLAQKIANEGYFFT VYVPYGNDWFAYFMRRLAERPQNLSLAIKE
FTKPKILKKVTLGIGIFATLLTSLILGIKRHHK*

Sequence 1961

40 Contig_0687_pos_8147_7746,
putative peptide of unknown function
atgttttttaacaaaatggctaccgaatggagacatttttgtgcaggattcctcttaatg
ttgaagtataacgcactacgcaaaattatgagttgtttgacatcggtatttaaatgcaacg
ggctctaaaaacacatctttctcaattagatgaggatgatcaattaattttaaacttagaa
aatagacaaaattgttcaacgtcatcctatgggtggcattcaaatatcccagaaacaaat
45 caagtcataagccccagatatggaaccttaaaaaatgtgattgcaattggacaaatgacc
aacggtgtcaataaaacttagaaatggcgtaaaagatgattgttaaatcaagttgttgataca
gtatctcaattatatataacacaggaataagataagtaa

Sequence 1962

50 MFLTKMATEWRHFCAGFLLMLKYNDTTQNYELFDIVINATGSKTHLSQLDEDDQLILNLE
NRQIVQRHPMGGIQIIPETNQVISPRYGLKNVIAIGQMTNGVNKLNRNGVKMIVNQVVD
VSQLYITQENRNK*

Sequence 1963

55 Contig_0687_pos_7297_6266,
is similar to (with p-value 3.0e-59)
>sp:sp|P17618|RIBG_BACSU RIBOFLAVIN-SPECIFIC DEAMINASE (EC 3
.5.4.-). >pir:pir|S45543|PN0100 ribG protein - Bacillus subt
ilis >gp:gp|L09228|BACDIA_10 Bacillus subtilis spoVA to serA

region. NID: g410114. >gp:gp|X51510|BSRIB_2 B.subtilis ribo
flavin biosynthesis operon ribG, ribB, ribA, ribH, and ribT
genes. NID: g40083. >gp:gp|Z99116|BSUB0013_40 Bacillus subti
lis complete genome (section 13 of 21): from 2395261 to 2613
730. NID: g2634723.
5 atggatgatgctattcaactagcaaaaatggtaaattggacaaacaggtgttaatccacca
gtaggatccgttgtgtgtaaaaaacggtaggattgtaggtttaggtgcacatttaaaaaag
ggagataaacatgccgaagtacaagctattgaaatggcaggtttaaataccaaggtgct
accatatacgtttcatttagaaccttgcacacaccatggttcaacaccaccttgtgtgcat
10 aaaaatcattgaagcgggcataatctaaggtcatctatgctgttaaagatactacttttagta
agtaaggggtgacgagattcttgagagaagctggtagaggttgaatttcaatataatgaa
aatgcagctgcattataccgtgacttttttactgctaaaagaaacgaagttccagaagta
actgtaaaggtctcatctagtctagatggtaaacaagcaacagactttaatgaaagtaag
tggaatacaaaacaaagaagttaaagaagatgtttatcaattaagacatgagcatgatgca
15 gttattactggcgtagaaccattgaagcagacaatccattgtatacaaccagggttcct
gatggaaagcatccgattcgagttattcttttctaagaaaggtcaactcgattttaatcaa
caaatatttaaaagatactgcacgagatatggatttactgaaaatgaaaaattaaaa
caaaataaaaagttttattaaaaataataaattagtaattgtgatcaacgacaatatta
caagacttatatcaaagaggtatttgggaaactgctagtcgaggcaggccaaatattaca
20 tctcaatttctccaatccaaacatctaaatgaactcattttatatatagccccgaaatta
attgggtggttctggcaaacatcaattttataagactgacgaggtcattgatttgctgaa
gcaactcaatttgaaattgttgattccaagttaattaatcaaaatttaaaattgaaatta
cgaaagaagtga

25 Sequence 1964
MDDAIQLAKMVNQGTGVNPPVGSVVVKNRIVGLGAHLKKGDKHAIEVQAIEMAGLNTQGA
TIYVSLEPCTHHGSTPPCVHKIIEAGISKVIYAVKDTTLVSKGDEILREAGIEVEFQYNE
NAAALYRDFFTAKRNEVPEVTVKVSSSLDQKQATDFNESKWITNKEVKEDVYQLRHEHDA
VITGRRTIEADNPLYTTRVPDGKHPHILSKKGLDFNQIIFKDTASEIWIYTENEKLLK
30 TNKSFIKIINISNCDTTTTILQDLYQRGIGKLLVEAGPNITSQFLQSKHLNELILYIAPKL
IGSGSKHQFYKTDEVIDLPEATQFEIVDSKLINQNLKLRKK*

Sequence 1965
Contig_0687_pos_6265_5627,
35 is similar to (with p-value 1.0e-44)
>sp:sp|P16440|RISA_BACSU RIBOFLAVIN SYNTHASE ALPHA CHAIN (EC
2.5.1.9). >pir:pir|S45544|A35711 riboflavin synthase (EC 2.
5.1.9) alpha chain - Bacillus subtilis >gp:gp|L09228|BACDIA_
11 Bacillus subtilis spoVA to serA region. NID: g410114. >gp
40 :gp|X51510|BSRIB_3 B.subtilis riboflavin biosynthesis operon
ribG, ribB, ribA, ribH, and ribT genes. NID: g40083. >gp:gp
|Z99116|BSUB0013_39 Bacillus subtilis complete genome (secti
on 13 of 21): from 2395261 to 2613730. NID: g2634723.
atgtctatgtttacaggtatcattgaagaaataggtactgtacaacaagttcgctctgaa
45 caatcagtaagaacgcttgaaattaaagcacaacatttttagttgatatgcatattggt
gattcaataaagtgttaacggtgcatgtttaactgtgatagatttcactgactcaagtttt
tcagttcaagtcatacaagggactgaaaacaaaacatatcttgggaagtgttcaacgtaat
acagaagttaatctcgaaagagccatgagtggaagtgggagatttgggtggacatttcgtg
ttaggtcatgttgatgagcttgaacaattttctaaaatcaatgaaactgctaactcaaaa
50 attattttctattaaaacaaactaaaaatattttgaatcaaatggtaaagcaaggttctata
actgtagacggaggttagtcttactgtattttgatttacctgattatacttttgatatacat
cttataccagaaacacgtcgatctactattctttcatctaaaaaagtgggcgacaaagtg
cacttgagctctgacgtactattcaaatatgttgaaaacatcatgaatcaaatcaatcg
cagttaacagaagaaaagcttagagcattttggttttag
55

Sequence 1966
MSMFTGIIIEIGTVQQVRSEQSVRTLEIKAQNILVDMHIGDSISVNGACLTVIDFTDSSF
SVQVIKGTENKTYLGSVQRNTEVNLERAMSGSRFGGHFVLGHVDELGTISKINETANSK
IISIKTTKNILNQMVKQGSITVDGVSLTVFDLHDYTFDIHLIPETRRSTILSSKKVGDKV

HLESDVLFKYVENIMNQNSQLTEEKLRAFGF*

Sequence 1967

Contig_0687_pos_2457_1804,

5 putative peptide of unknown function
 atgtggaagtgggaaacagaaaatgacgcaaaaggcgttgttgcattgctcataatatt
 ttgaacatacaggtagatatgcatatgttatcacgatgttaagacgaaatgggtatcac
 gttatcatggcgatttaccgggacaaggcacaacttcacgagctcaaaagggacaaata
 gatgatttttaatacgtatcatgaaaatatattagagtggataaaaatagctaataatgaat
 10 aaaattccaacatttgttttaggtgtgggactaggtggtctcatcttttaaatctgctt
 gagaaaacagaattacatttgagggtatcttgttatcttccacctatgttagaactaaag
 agagactataaaaggcgcaaaaataaattgatttctaatgttggtaaaatttctaaagat
 actagattttaaagttggtataaactcctcaagatttaacacgtaatgatgaaattattgaa
 gaaacagcaaatgatggactaatgcttaaaaaggtaacatatagttggtataaccttata
 15 aatgaaaagatgaaagaaacaatggatcatatcagagatattaacctatttcagcattg
 ataattgtatggtaccaatgataaaaatttttcaacaattcattttgtatgattag

Sequence 1968

20 MWKWETENDAKGVVVIHNIHLEHTGRYAYVITMLRRNGYHVIMGDLPGQGQTSRAQKGQI
 DDFNTYHENILEWIKIANEYKIPTFVLGVGLGGLIILNLEKTELPIEGILLFSPMLELK
 RDYKGRKNKLISNVGKISKDTRFKVGITPQDLTRNDEIIEETANDGLMLKKVTYSWYNLI
 NEKMKETMDHIRDIPISALIMYGTNDKIFQQFILD*

Sequence 1969

25 Contig_0687_pos_1802_1392,
 putative peptide of unknown function
 gtgtaccctaataatataaattgagtttaaaataaaaaacagtaaaattaagatggaaattt
 catatccttgccttcttatcactataaatatttgaataaaagggtgatttttttatatac
 gatattttattattgaattataaaaaaatatatgcgataggtgttactataaataaactc
 30 aatacagttctctcaaacatataaccactaaatctcttatgcggatgggtatatcagta
 gagagaatactaggcacactagctgaaaagaatataattgtagaaacagaagtaacagca
 attataaatttagtgattaatgggtgcttcgactacaattaatgaaggcaaaaacatttct
 gttataccaatagctgcgccttttgcagttaaaaaggaatctggtatttga

Sequence 1970

35 VYPKYIIIEFKIKNSKLWKPHILAFILITINICKKGDFFIYDILLNKKIYAIGVTIILK
 NTVLSNHITTKSLMRMGISVERILGTLAEKNIIVETEVTAINLVINGASTTINEGKNIS
 VIPIAAPFAVKKESGI*

Sequence 1971

40 Contig_0688_pos_4702_3980,
 is similar to (with p-value 7.0e-53)
 >sp:sp|P42312|YXJA_BACSU HYPOTHETICAL 43.7 KD PROTEIN IN KAT
 B 3'REGION. >gp:gp|Z99123|BSUB0020_197 Bacillus subtilis com
 45 plete genome (section 20 of 21): from 3798401 to 4010550. NI
 D: g2636240. >gp:gp|Z99124|BSUB0021_6 Bacillus subtilis comp
 lete genome (section 21 of 21): from 3999281 to 4214814. NID
 : g2636442. >gp:gp|D83026|D83026_20 Bacillus subtilis genome
 sequence covering lic-cel region. NID: g1783231.
 50 atgatgtctatgagttcagtttcaggagcaattgtggcgcttatgtgcaaatgatacct
 ggagaacttgtattgacggcaattccacttaattatttaacgcaattatagtttcttgt
 attttgaatcctgtatcagttgaagaacaagaagatgtcgtgtatagcattaaagatcac
 caaactgaaagacaaccatttttctcatttcttggagattcagttttagcagctggaaag
 cttgtatttaattatcatttgcatttgcattagctttgtagctttggctgacttaattgat
 55 agattgattcatttaatacacacatcttattgcaaatggtattggtgtcaaaggtagcttt
 ggtcttgatcaaactcttaggcgttttcatgtatccatttgcctttactattaggtttaccg
 ttttaatgaagcgtgggaagtagcacaacaatggcgaagaaaattgtaacaaacgaattt
 gttgtgatgggggaaatttctaatcaagtcaatgcgatgacgcctcatcatagagcagtt
 atatcaacatttttagtttcttttgcaaacctttcaactatttgaatgattataggtaca

ttgaaaggtattgttgataagaaaacgtcggatttcggttccaaatatgtaccgatgatg
 ttgtagcaggaatttttagtatccttacttactgctgcatttggttgattatttgcttg
 taa

5 Sequence 1972
 MMSMSSVSGAIVGAYVQMIPGELVLTAIPLNIINAIIVSCILNPVSVEEQEDVVYSIKDH
 QTERQPFSSFLGDSVLAAGKLVLIIFAFVISFVALADLIDRLIHLITHLIANGIGVKGSF
 GLDQILGVFMYPFALLLGLPFNEAWEVAQQMAKKIVTNEFVVMGEISNQVNAMTPHHRV
 ISTFLVSFANFSTIGMIIGTLKGIVDKKTSDFVSKYVPMMLLAGILVSLTAAAFVGLFAW
 10 *

Sequence 1973
 Contig_0688_pos_3410_2610,
 is similar to (with p-value 2.0e-26)
 15 >sp:sp|P54478|YQFU_BACSU HYPOTHETICAL 32.5 KD PROTEIN IN CCC
 A-SODA INTERGENIC REGION. >gp:gp|D84432|BACJH642_146 Bacillu
 s subtilis DNA, 283 Kb region containing skin element. NID:
 g2627063. >gp:gp|Z99116|BSUB0013_221 Bacillus subtilis compl
 ete genome (section 13 of 21): from 2395261 to 2613730. NID:
 20 g2634723.
 atgataggttcatttattttctctgcaggtatcaatgcatttggtatttcagggaatttg
 ggtgaggggtggtgtcactggtatagccatcgattatattatgcttttcatttcaccg
 ggaataaccaatttcggttttaaatgctattttaattattgtgggtataaatatttgagt
 aaacgtagtacatatttaacaatttttgctacagtactcatttcaatctttctaggttta
 25 actgaaacatggcatgtagaaactgggaatgttggtgattaatgctgtgttcggtgggact
 tgtgttggttaggaattggtattatcggttttagcagggggaacaaccgctggaacggtt
 attcttgcgagaattgttaataaatatttagatattagtagcgccttacgctttgttattc
 tttgaccttatcggttggtgcttatttcattgacagaaattccttttagtgaagtgttagtt
 acagttatgtctttatatataggtacaaaagtgtggaatttggtatagaaggattaaat
 30 actaaaaaggaatgactatttatctagtcgccctaataaggttagcaaaagctattgat
 cagcaagttggaagaggattaacaatatataatggacacggttattacactagagaagaa
 aaagatgtactttacgtagtcactctctaaaaacacaagtatctcgtgctaaacgaatcatc
 aaaaatattgacgaaaatgccttttagttattcatgacgttcgtgatgtatatggtaat
 ggttttttggtagatgagtaa

35 Sequence 1974
 MIGSFIFSAGINAFVISGNLGEQGVGTGIAIVLYYAFHISPGITNFVLNAILIIVGYKYL
 KRSTYLTIFATVLISIFLGLTETWHVETGNVNAVFGGTCVGLGIGIIVLAGGTTAGTV
 ILARIVNKYLDISTPYALLFFDLIVVLISLITEIPLVKCLVTVMISLYIGTKVMEFVIEGLN
 40 TTKAMTIISSRPNEVAKAIDQQVGRGLTILNGHYTREETKDVLYVVISKTQVSRAKRII
 KNIDENAFVLDVDRDVGNGFLLDE*

Sequence 1975
 Contig_0688_pos_2251_1454,
 is similar to (with p-value 2.0e-89)
 45 >sp:sp|P49938|FHUC_BACSU FERRICHROME TRANSPORT ATP-BINDING P
 ROTEIN FHUC. >gp:gp|X93092|BSFHUDBG_4 B.subtilis fhuDBG ge
 nes. NID: g1070011. >gp:gp|Z99121|BSUB0018_16 Bacillus subti
 lis complete genome (section 18 of 21): from 3399551 to 3609
 50 060. NID: g2635827. >gp:gp|AJ223978|BS43KBDNA_12 Bacillus su
 btillis 42.7kb DNA fragment from yvsA to yvqA. NID: g2832786.
 atgagtcgtttaagtgggtgaacaagtgaataattgggtacggtgattctacgattattaat
 aatttgatgtcgcaattcctgatggaaaggttacttctattattggacctaacgggtgt
 gggaaatcaactttatttgaagcggttatctagactattgtcaattaaagaaggtaaaatt
 55 aatttgatgggtaagagtattcatgccacatccacgaaagaaatagctaaaaaaatagca
 attttaccacaatcaccagaggtcccagatggacttactgtaggagaacttglttcttat
 gggcggtttccacatcaaaaaggatttggtcggttaactgcagaagataaaaaagaaatt
 gattgggcattgtcagttacaggtacaagtgaatttcgtcatcgtactataaatgattta
 agtgggtggacaaagacaacgcgtgtggattgcaatggcactagcccaacgtactgatatt

atcttcttagatgaacctacaacttatttagatatttgtcatcaattagaaatattaaat
 ttagtcaaaaagctcaacgaagaagaagggttgactattgtgatgggtttacatgacatt
 aatcaagcaattcgcttctcagatcatctcattacgatgaaagctggagatattgttgct
 actgggtcaaaactgatgaagtgttaactaaggacattttagaaaaggatttaattattgat
 5 ggtggttttagatatagatccgagaacagggaaaccaatttttagttacttacgatttattc
 tgtcagacgtatttcgtga

Sequence 1976

MSRLSGEQVKIGYGDSTIINNLDVAIPDGKVTISIIGPNGCGKSTLLKALSRLLSIKEGKI
 10 NLDGKSIHATSTKEIAKKIAILPQSPEVPDGLTVGELVSYGRFPHQKGFGRLLTAEDKKEI
 DWALSVTGTSEFRHRTINDLSGGQRQRVWIAMALAQRDIIIFLDEPTYLDICHQLEILN
 LVKKLNEEEGCTIVMVLHDINQAIRFSDHLITMKAGDIVATGQTDEVLTKDILEKVFNIID
 GVLDIDPRTGKPILVTYDLFCQYS*

15 Sequence 1977

Contig_0688_pos_1134_634,

putative peptide of unknown function

atgcaacatcttataaaaaaacatgtattgaatggcgagtttgaactagttagacagtta
 atgtccgaaacagatttttatggaatttgaagaagcatacatctctagtgtcatgaagta
 20 gagagtatgatgttttatacatgtattctagatatgattaaggtagaagaatcatcagaa
 ttacatgatttagcatttcttttacttgtttatcctttaagtgaatatgagggcgacta
 gattcagcttattatcatgcagattcttccataaaacttactgacggaaatgaagtgaag
 agttttattacaaatgttattgttcatgtctattcctgagccagttatttcggataaaaaa
 gcggttgatgtcgctaaacgaattctaaaactcgatccaagtaataatgttgcacgaaat
 25 gtacttaagatacagcaaaacgtatggataatgtagttgtagacattaatgaattgaac
 aatcaaagagatgcacgctaa

Sequence 1978

MQHLLIKKHVLNGEFELVRLMSETDFMEFEEAYISSAHEVESMMFYTCILDMIKVEESSE
 30 LHDLAFLLLVYPLSEYEGALDSAYYHADSSIKLTDGNEVKSLQLMLLHAIPEPVISDKK
 AFDVAKRILKLDPSNNVARNVLKDTAKRMDNVVVDINELNNQRDAR*

Sequence 1979

Contig_0688_pos_0_409,

35 putative peptide of unknown function

atggggctgaataaagaagctataaaaattggttttgcctatgtcggcattgttgtcggc
 gcaggattttcaacaggacaagaagtgatgcaatttttcacaccatttggttatggtca
 tatattggagtgattatctcaggatttatacttgattcataggaagacaagtagctaag
 ataggtactgcatttgaagcgaaaaaatcacagagtcacattgcaatatgtgtttggaaaa
 40 aaatttagtaaagtttttgattatattcttcttcttttcttatttggtatagctgtcact
 atgatagccggatcaggttctacttttgagcaaagtttggaaattcctacttggttaggc
 gcattaatcatgacagttttgatttacttaacattattatttagCGTCAA

Sequence 1980

MGLNKEAIKIGFAYVGIVGAGFSTGQEVMQFFTPFGLWSYIGVIISGFILGFIGRQVAK
 45 IGTAFAEAKNHSTLQYVFGKKFSKVFDYIILVFFLFGIAVTMIAGSGSTFEQSFGIPTWLG
 ALIMTVLIYLTLLLASX

Sequence 1981

50 Contig_0690_pos_3925_4464,

is similar to (with p-value 7.0e-61)

>gp|Z99107|BSUB0004_107 Bacillus subtilis complete genome
 (section 4 of 21): from 600701 to 813890. NID: g2632866. >g
 p:gp|Y15254|BSYERABCD_4 Bacillus subtilis 13kB DNA fragment,
 55 from yerA to sapB gene. NID: g2577959.

atggtagagatagacacctatccaagctttattactgttgatgggtggtgaaggtggtaca
 ggcgctaccttccaagagcttgaagatggtgttggtttaccgttatttacagcacttct
 atcgtttcaagtatgttagaaaagtatggcataaagaacaagggttaaaatttttgcgtcc
 ggtaaaattagtgactccagataaaatcgcaattgcattaggattaggtgcggatctcgtc

aatattgctagaggtatgatgataagtgtaggatgcatcatgagtcacaatgtcattta
aatacatgtccagttggagtagcaacaaccgatcctaaaaaagaaaagggacttattgtt
gatgaaaaacaataaccgtgttacaattatgttacaagtttgcaggaaggtttatttaac
atcgctgcagctgtagggtgttcatagtccaacggagattacttccgaccatattatctat
5 agacaatttagatggcactacaacgctccattcaggattataaaacttaaattaatttcttaa

Sequence 1982

MVEIDTYPSEFITVDGEGGTGATFQLEDGVGLPLFTALPIVSSMLEKYGIRNKVKIFAS
10 GKLVTDPKIAIALGLGADLVNIARGMMISVGCIMSQQCHLNTCPVGVATTPKKEKGLIV
DEKQYRVNTYVTSLHEGLFNIAAAVGVHSPTEITSDHIIYRQLDGTTSIQDYKCLKLIS*

Sequence 1983

15 Contig_0690_pos_3157_2813,
putative peptide of unknown function
gtgacaaaccggaggaaggtggggatgacgtcaaatacatcatgcccttatgatttgggc
tacacacgtgtcataatggacaatacaaaaggcagcgaaaccgcgaggtcaagcaaacc
cataaagttgttctcagttcggattgtagtctgcaactcgactatatgaagctggaatcg
20 ctagtaatcgtagatcagcatgctacgggtgaatacgttcccggtcttgtaacacccgcc
cgtaacacccgagaggtttgtaacacccgaagccggtggagtaaccatttggagctaacc
gtcgaaggtgggacaaatgattgggggtgaagtcgtaacaaggtag

Sequence 1984

25 VTNRKVGMTSNHHAPYDLGYTRATMDNTKGSETARSSKSHKVVLSDDCSLQLDYMKLES
LVIVDQHATVNTFPLVHTARHTTRVCNTRSRSNHLELAVEGGTNDWGEVVTR*

Sequence 1985

30 Contig_0690_pos_0_685,
is similar to (with p-value 5.0e-96)
>gp:gp|AF029225|AF029225_1 Staphylococcus carnosus NarG, Nar
H, NarJ, and NarI genes, complete cds. NID: g3929521.
atgaagcacttggttaggtgcgcgctctggtttaatggcagagccaaatgaagatgataaa
ccagagggaaattaaatggcgcgaggatacagaagggaaacttgatttattagatcactt
35 gatttcagaatgactgcgacgccattatattcagatatcgttttacctgctgcaacttgg
tatgaaaaacatgatttatcttctacagacatgcattcatttcatccatttaaccca
gcatgtgacccattatgggaatcgcggttcggactgggatatttataaaactctaagtaaa
gctgtttcagaaatggcgaagattatcttccaggtaaatttaaagatgtcgtaactaca
ccattagagcatgattcacaacaagaaatttcaactgaatacggatttgtaaaaagattgg
40 tctaaaggagaaattgaaggtgtgccaggtaaaacaatgcctaatttttctatcgtagag
cgagactatacacaatttacgataaattcggtactgttggtccaaaactagaaaaagg
aaaataggtgctcatggtgtgagttatagcggttagtgaagagtacgaagaactaaaagt
atagttggaacttggaatgatgataatactatttcagttaaaaatgatagaccgagaata
gatacagcgagaaaagtagcagaGG

45

Sequence 1986

MKHLGARSGLMAEPNEDDKPEEIKWREDTEGKLDLLVSLDFRMTATPLYSDIVLPAATW
YEKHLSSDTMHPFIHPFNPAIDPLWESRSDWDIYKTLKAVSEMAKDYLPGKFKDVVTT
PLGHDSKQEISTEYGIVKDWKSGEIEGVPGKTMFNFSIVERDYTIYDKFVTGPKLEKG
50 KIGAHSVSYSVSEYEELKSIVGTWNDDNTISVKNDPRIDTARKVAEX

Sequence 1987

55 Contig_0691_pos_3686_4519,
is similar to (with p-value 1.0e-59)
>gp:gp|U40604|LMU40604_6 Listeria monocytogenes ClpC ATPase
(mec) gene, complete cds. NID: g1314293.
atgagacgtagtgcggtagaaatattatttgcataattggtttaattattggtttattt
atttcagtgatggtttcttttatcttagaaatgataggttaattccatattaaatcacttt
gtacctatgataatcactattattttatgttatttagggtttcaatttggtctgaaaaaa

agagatgaaatgcttatgtttttaccagagaatatggcacgttccatgtctaataatata
cgaagagcgacacctaagattgtagatacaagtgccattatcgatggaaggatattagat
attatacgttgccgatttatcgatggtgatataattgataccacaaggcgttataaatgaa
ttacaggttatagcggatgctaaagatagcgtgaaacgtgaaaaagggtcaaagaggatta
5 gatattttgaatcaactttatgatttagattatcctacacgcgttatatccaactcaa
tcccatagtgatatagatacattattaattaaattagcacaacagtatcatgcacatgtg
attacgactgattttaatttaaataaagtatgtcacgttcaaggaattacagcactcaac
gttaatgatttatcggaagcaatcaaacctaatgtacatcaaggcgaccagttaagtatt
ttattaacgaagataggtaaagagccaggacaaggcgtaggatatttagatgatggtaca
10 atggtggttgttgataacgcgaagagttacattggtcaacaagttaatttagaggtcgta
agtttgttacaacatcatcaggaagaattgttttgcgaaatttgttgactga

Sequence 1988

MRRSAVEILFATIGLIIGLFISVMVSFILEMIGNSILNHFVPMIITIILCYLGFQFGLKK
15 RDEMLMFLPENMARSMNNIRATPKIVDTSIIDGRILDIIRCGFIDGDILIPQGVINE
LQVIADAKDSVKREKQQRGLDILNQLYDLDPTRVIHPTQSHSDIDTLLIKLAQQYHAHV
ITDFNLNKVCHVQGITALNVNDLSEAIKPNVHQGDQLSILLTKIGKEPGQGVYLDGDT
MVVVDNAKSYIGQQVNLEVVSLQTSSGRIVFAKFVD*

20 Sequence 1989

Contig_0692_pos_2251_3219,

putative peptide of unknown function

gtgggtagggaaattccaccacatacttataaatttaaagatcaagaaacatatgaaagc
tgataaggaatttagctcttcatcaaggtaaaaaatctactttcaatatattcatgat
25 taagataattttacctaagaatattatgcacttgataaagatgttttgttgctctta
aataaagcacgaattccagaatggactaataacaaatatctaccacaaagagaaattgtc
tcaattaaagattttgaatctcatattcaagcatggtcgtatccatttgcataaaacca
ggcgtatgatttacctacagcaggaggatattggtgttatgatttgtataatgatagat
ttagctaaagctatcacacgcatacaaatgcatacagagactgaaaatttaattcatt
30 gaacaaaaaattaatgcagtgaataactattgtgtacaatttgcattatcagatgatatt
ggtatcaaaacttaggaacagcgcaacagttaactaatgactatggattttacaacgga
aatgataatttttaactgattgtgcctcagaatgtaataagacgctggtagagagatttgaa
ataggcgtaagcaaaaggttttttgggtgtagcaggttttgacttactagtagatgataat
aatgatgtttatgcgattgatttaaacttttaggcaaaacggatcaacgagtatgctactt
35 ttagcaaaagatttaactcatggatatcataaattttacagttacttttctaattggagat
aatacaaaattctataatgctattttaaaatacgtagaattaggtgtactttatccactt
tcctattacgatggagattggtatggaaagaatcaagttaattctagatttggctgcatt
tggcatggggaaaaataagaattaattaatcgatatgaacaacaatttatattggaagct
ggattataa

40

Sequence 1990

VGREIPPHYTKFKDQETYESLIRNLALHQGKKIYFQYIHDEDILPKEYYALDKDVFVALN
NKARIPETWNNKYLPRQREIVSIKDFESHIQAWSYPFVIKPGDDLPTAGGYGVMICYNDTD
LAKAITRINNASAETENLIEQKINAVNNYCVQFAYSDDIGIKYLGTAQQLTNDYGFYNG
45 NENVNDVPQNVIDAGREIMEIGVSKGFFGVAGFDLLVDDNNDVYIDLNFRQNGSTSMML
LAKDLTHGYHKFYFSNGDNTKFYNAILKYVELGVLYPLSYDGDWYGKNQVNSRFGCI
WHGENKELINRYEQQFILEAGL*

Sequence 1991

50 Contig_0692_pos_3392_4057,

putative peptide of unknown function

atgtcatacaaatgaagcattttttaagatatatttgattaatgaatatatttatttt
gcttcaaaaaataaaaaattaattagaatacaacatgagaatttgccatatattgctatg
55 tggacagacgaaaatgttgctgagtccttatttgttacatcattcaattgattacgacaaa
atcatttagagcagatatattgaccgttttgaacatatgaaatggatgaaatctttgatcca
ggtgacaaaagtttttagttaatgtgaataatggtgaagaaggaaacattgtagatatagtt
aaaatgactgatgagttgatgtctgaattagatgatataagaatgagagagtttattaaa
gatgtcgcaaaatatgacgaagtatacggattgacaaacaaagggtgaaaagaattttatt
atgatttcagatgatgaccataacaaaccacacatcatgcctgtttggagtattaagagt

agagcgcgtaaagtacgtgatcaagatTTTTgaagaatgtgatttaatcgaaattgaaggt
gaagtctttagtgaaatggttagacaagttacgcgatgataataaagcagtagcgattgat
ttgaaatcaggtgttgttggtactgttgtatcagcgcacaaaactgtcaaatgaagcaaca
ttttaa

5

Sequence 1992

MSYKYEAFFKDILINEYIYFASKNKKLIRIQHENLPYIAMWTDENVAESYLLHHSIDYDK
IIRADIDRFVYEMDEIFDPGDKVLVNVNNGEENIVDIVKMTDELMSELDDIRMREFIK
DVAKYDEVYGLTNKGEKNFIMISDDDHNKPHIMPVWSIKSRARKVRDQDFEEDLIEIEG
EVFSEWLDKLRDDNKAVAIDLKSGVVGTTVVSAQKLSNEATF*

10

Sequence 1993

Contig_0692_pos_4711_4280,
putative peptide of unknown function

15

atgatacaaggttttaggctatttattgtccaataataacagattataaagaattaacgaat
ttagctcaaaatggagatcgtgatgccattgatttaaaagtaaaacatatttataaagat
actgaaccaccaatttcctggagatttaacagcagcaaatTTTggaaatgtattacatcac
ttagataatcagtttacatcagctaacaaacttgccctctgcaattggcgtcgttggtgaa
gttataacaactatggctattacattagcacgtgaatataagactaagcacgttgatatat
atcggttcatcatttaataacaatacaattactacgtgaagttgttgaataattacactgtt
ctaagaggatttaaacctgactatattgagaatgggtgctttttcaggcgctttaggagca
ctttacctctaa

20

Sequence 1994

25

MIQGLGYLLSNITDYKELTNLAQNGDRDAIDLKVKHIYKDTEPPIPGDLTAANFGNVLHH
LDNQFTSANKLASAIGVVGEVITTMAITLAREYKTKHVYIGSSFNNQLLREVVENYTV
LRGFKPYYIENGAFSGALGALYL*

Sequence 1995

30

Contig_0693_pos_7050_7454,
putative peptide of unknown function

35

atgataaaagatggtattggagcaatagctcctcttgggagtggggaaacttatggctac
catactttagatcaacatatacaagattaccctcataatgtaactagatttttagtagtg
aaaaatcatactcattttattgaacatccaaacacaactatcttccttatcacgcctaag
tatgataagccaggacttttagctagtgttttaataacttttactttattcaacataaat
ttatcgtggattgaatctagaccacttaaaactcaattaggtatgtatcatttttatgtt
caagccgatactgctataaataatgatgtgaataaaatttttcaatttttagagactttg
gattttcaagttaaaattatcggcgcttttaataagaaaaactaa

40

Sequence 1996

MIKDGIGAIAPLGSGETYGYHTLDQHIQDYPHNVTRFLVVKNHTHFIEHPNTTIFLITPK
YDKPGLLASVLNTFTLFNINLSWIESRPLKTQLGMYHFYVQADTAIINNDVNKIISILETL
DFQVKIIGAFAFNKKN*

45

Sequence 1997

Contig_0693_pos_7551_8120,
putative peptide of unknown function

50

atgagtttttgtaaattcatttactggcttctttatttacaatactttgtattgttggt
tgctttcttatgagacagcttcagcctataaaagggaatgacaaaaatttaaacgtaact
agcaaggtgaacaatcaacgttgggttaattacagagcgtcaaacaggtccacatattaat
tttagaatacaaggtaaagttagtaatcacattacgattacagtaacctaaatatattaaa
aacatagatatataaaactaatgccggggatttaaatattgttggagtaaatagtggcaca
ggaagatttgatgctgaatctggagacattaaagttcaaaaaggacgatataaaaagggtg
acacttcataatgaggatggggatattcaaatgaaacaattagaccctgatattccttta
cgtattataaaatgaagaaggggatataaactgaattataaaaaagaacttcacacacc
caaatcatcactcgtaatgaagaaggggaaacagacatcgatcatcgtgtgttatataat
agtaaagtactatttagtgacgtttcataa

55

Sequence 1998

MSFCKFIYWLLYLQYFVLLACFLMRQLQPIKGNOKNLNVTSKVNNQRWVITERQTS PHIN
FRIQGVKSNHITITVPKYIKNIDIKTNAGDLNIVGVNSGTGRFDAESGDIKVQKGRYKKV
TLHNEDGDIQMKQLDPDIPLRIKNEEGDINLNYKKELHHTQIITRNEEGETDIDHRVLYN
SKVLFSDVS*

5

Sequence 1999

Contig_0693_pos_9026_9604,

is similar to (with p-value 4.0e-47)

10 >sp:sp|P42085|XPT_BACSU XANTHINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.-). >pir:pir|S51309|S51309 xanthine phosphoribosyltransferase - Bacillus subtilis >gp:gp|L77246|BACYACA_2 Bacillus subtilis (YAC10-9 clone) DNA region between the serA and kdg loci. NID: g1256615. >gp:gp|Z99115|BSUB0012_148 Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220. NID: g2634478. >gp:gp|X83878|BSXPTPBUX_1 B. subtilis xpt and pbuX genes. NID: g633168.

15 gtggagtcgttaggacgaaaagtcaaagaagatggcgttgcctcatcgatgagaaaattttg
aaggtagatggatttttaaatcatcaaattgatgcaaagttgatgaatgatgtaggttaa
acatttttagtgcctttcaaagacgctggtattactaaaattttaactattgaagcttct
20 ggtattgcgcctgctattatggcttcttttcattttgatgttccttgcctatttgctaaa
aaagctaaacctagtagtctttgaaagatggcttttatagcacggatattcattcatttaca
aaaaataaaacgagtagcagtcattgtatctgaagaatttttaggtgcagacgataaagta
cttatcattgatgacttttagtcaatggtgatgcttcgctaggtcctaatgacattgta
aaacaagcaaatgagacgacagttggcgtgggtattgtggttgaaaaaagtttccaaaat
25 ggtcgccaacggttagaagatgcaggcttatatgtatcttcactttgtaaggtagcttca
ttaaaggcaataaggtaactcttttaggtgaagcgtaa

Sequence 2000

30 VESLGRKVKEDGVVIDEKILKVDGFLNHQIDAKLMNDVGKTFYESFKDAGITKILTIEAS
GIAPAIMASFHFDPCLFAKKAKPSTLKDGFYSTDIHSFTKNKTSTVIVSEEFLGADDDKV
LIIDDFLANGDASLGLNDIVQANATTVGVGIVVEKSFQNGRQRLEDAGLYVSSLCKVAS
LKGNKVTLLEA*

Sequence 2001

35 Contig_0693_pos_9643_10872,

is similar to (with p-value 8.0e-95)

>sp:sp|P42086|PBUX_BACSU XANTHINE PERMEASE. >pir:pir|S51310|S51310 xanthine permease - Bacillus subtilis >gp:gp|L77246|BACYACA_3 Bacillus subtilis (YAC10-9 clone) DNA region between the serA and kdg loci. NID: g1256615. >gp:gp|Z99115|BSUB0012_147 Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220. NID: g2634478. >gp:gp|X83878|BSXPTPBUX_2 B. subtilis xpt and pbuX genes. NID: g633168.

45 atgtatgcaggggctattcttgcctattattgtggggacaaagcttaaaattttcagct
gaagaaattgcttatctagttactgttgatattttatgtgcgggtagcgacatttctt
caagcaataaaagtcacaggactggattaccgattgtactaggatgtacgtttactgcc
gttgccactatgatactcatcggtcaaagcgaaggacttgatgttttatatggttcgctt
ttaatatccggtatcttagttgttttaattgcacctttttctcttatttagttaaatc
tttccacctgttgtaacaggaagtgttgtagacaattattggaatcaatttaagccagtt
50 gcaatgaattacttggcaggtggtgaaggagcgaaaaactatggcgatactaagaattta
atattagggtggtgttacactactcattattcttattttgcaaagatttacaagggttc
ttgaaatcaattgcgatacttataggattagcaataggtagctttagctggtatattt
ggaatggttgatatacaacaagtgggtgatgcacattggtttggtttccctgtgccattc
agattttctggcttcggatttgatgtcagctcaatacttgatatttttcattgttgagtt
55 gtaagtttaattgaatctactggtgtctatcatgcactgagtgaaattactggtagaaaa
ctagaaagaaaagattttcgaaaagggtacactgcggaaggcttagcaatcattttagggt
tcaatatttaatgcgtttcccttacactgcataatcccaaaatgtaggtcttggtttctta
tcaggagctaaaaagaacaatgtgatatatggaatggttattcttttactaatttgcggt
tgtatacctaaattagggtgctttagctaatattattccattgccgggttttaggtggagca

atgatagcaatgtttggaatggttatggcatacggcggttagtattttgggtaacattaat
 ttccaaaaataaaataattttattaattattgcaatttcagtagggtaggtgctggtatt
 agtgacgtacctaagcatttaaaggattaggagaacaatttgcttggttaactcaaaat
 ggtatagtgttgccgcaatttctgcaatcatcttaaatttctttttaatggtataaag
 5 tataaacaactgaagaaatgtgaaataa

Sequence 2002

MYAGAILVPIIVGTSLSKFSAEIAYLVTVDFMCGVATFLQANKVTGTGLPIVLGCTFTA
 VAPMILIGQTKGLDVLGSLISGILVVLIAPFFSYLVKFFPPVVTGSSVVTIIGINLMPV
 10 AMNYLAGGEGAKNYGDTKNLILGGVTLLIILILQRFTKGFLKSIAILIGLAIGTALAGIF
 GMVDIKQVGDAHWFGFPVPRFSGFGFDVSSILVFFIVAVVSLIESTGVYHALSEITGRK
 LERKDFRKGTYAEGLAIIIGSIFNAFPYTAYSQNVGLVSLSGAKKNNVIYGMVILLICG
 CIPKLGALANIIPLPVLGGAMIAMFGMVMAYGVSILGNINFQNNLLIIAISVGLGAGI
 SAVPQAFKGLGEQFAWLTQNGIVLGAISAILNFFENGIKYKQTEENVK*

Sequence 2003

Contig_0693_pos_10910_11434,
 is similar to (with p-value 2.0e-63)
 >sp:sp|P21879|IMDH_BACSU INOSINE-5'-MONOPHOSPHATE DEHYDROGEN
 20 ASE (EC 1.1.1.205) (IMP DEHYDROGENASE) (IMPDH) (IMPD). >pir:
 pir|S12623|DEBSMP IMP dehydrogenase (EC 1.1.1.205) - Bacillu
 s subtilis >gp:gp|X55669|BSIMPDE_1 Bacillus subtilis guaB ge
 ne for IMP dehydrogenase. NID: g39958.
 atgtgggaaaataaatttgcataagaatctttaacattcgacgacgtgttactcattcca
 25 gctgcatcagatgttttaccaagcgatgttgacttaagtgtcaaatatcagataagatc
 aagttaaacattcctgttatctcagcaggtatggatacagtaactgaatcaaaaatggca
 attgctatggctcgacaaggcggttttaggtgttattcataagaatattggcgctcgaagag
 caagctgatgaggtacaaaagggttaaacgttcagaaaatggtgttatttctaaccgttc
 ttcttaacaccggaagaaagtgtgtatgaggctgaagcattaatgggtaaataaccgtatc
 30 tctggtgtacccattgtcgataatcaagaggatcgcaagttgattgggattttaacaaat
 cgtgatttacgttttattgaagatttttcaattatcatagtcattttctattaattcc
 tctaaagcattccaattgttttggccttcttttgacatagagtga

Sequence 2004

MWENKFAKESLTFDDVLLIPAASDVLPDVLSDVLSVKLSKIKLNIPVISAGMDTVTESKMA
 35 IAMARQGLGVIHKNMGVEEQADEVQVKRSENGVISNPFLLTPESVYEAALMGKYRI
 SGPVPIVDNQEDRKILIGILTNRLRFIEDFSIIIVNISINSSKGIPIVLASFIDIE*

Sequence 2005

Contig_0693_pos_17655_16912,
 is similar to (with p-value 3.0e-97)
 >sp:sp|P50849|PNPA_BACSU POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANS
 FERASE (EC 2.7.7.8) (POLYNUCLEOTIDE PHOSPHORYLASE) (PNPASE).
 >gp:gp|Z99112|BSUB0009_139 Bacillus subtilis complete genom
 45 e (section 9 of 21): from 1598421 to 1807200. NID: g2633902.
 >gp:gp|U29668|BSU29668_2 Bacillus subtilis ribosomal protei
 n RpsO (rpsO) gene, partial cds, and polynucleotide phosphor
 ylase (pnpA) gene, complete cds. NID: g1184678.
 atggatgccgggtgtaccaattaaagcgccagtcgcagggattgcaatgggactagtaacg
 50 cgtgacgatagctatacaattttaactgatattcaaggaatggaagatgcattaggtgat
 atggacttcaaagtagcaggtactaaagacggtattactgcgattcaaattggatattaaa
 attgatggtttaactcgagaagttattgaagaagcactagaacaagcgcgtaaggacga
 ttagctattatggatcatatgcttcacacgattgaacaaccacgcgaagaattaagtgtc
 tacgcacaaaagtggtaactatgagttatattccagataaaaattcgagacgtgattgga
 55 ccaggtggtaagaaaatcaatgaaattatcgacgaaactggagttaaattagatattgaa
 caagatggtacaattctttataggtgctgtagatcaagcgatgattaaccgtgcaaaagaa
 attatcgaaagatattacacgcgaagcggaagttggacaagtatatcatgctaaagtaaaa
 cgtattgaaaagtatggtgctttcgttgattgttccttggtaaagacgcgttattacac
 atttctcaattttcacaagaaagaattaataaagtagaagatgttcttaaaattggagat

acaattgaagtgaaaattactgaaatcgataaacaaggctcgcgttaatgcgtcacataaa
gtattagagcaatctaaaaattaa

Sequence 2006

5 MDAGVPIKAPVAGIAMGLVTRDSYTIILTDIQGMEDALGDMDFKVAGTKDGITAIQMDIK
IDGLTREVIEWEEALEQARQGRRLAIMDHMLHTIEQPREELSAYAPKVVTMSINPKIRDVIG
PGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAKEIIEDITREAEVGVYHAKVK
RIEKYGAFVELFPGKDALLHISQISQERINKVEDVLKIGDTIEVKITEIDKQGRVNASHK
VLEQSKN*

10

Sequence 2007

Contig_0693_pos_15366_14407,

putative peptide of unknown function

15 atgaaagacaacaaacctaataattcgaaattaattcaaacatattttaagtaagaaaact
ttaagatatggtacagcaagtgcattaacattggcactctattttatttaacagtaacgta
actgtgtatgcggatgaaaatactgcaaaccaaaatcaaggaacatcaccaaaaacttca
cagacagcacctacaaataatactgaaaatacagatgccacagccataacaacagatcaa
aataataatgatgaagaagaatacagatgcgtcatatgaacttccaattctttatgtaact
gtctggctagatgatcaaggaaatattttaagatgctgtggaagatgctaaaaccct
20 gcttcagaaaaggcaaccggtgaaaattcctgggtaccaacattatagaacttctgtgagt
gacggaattactaagttttatttatcgtaaaattagcactgcacaatcacctatagttgaa
aataatcaacaagataataatacaataaagttgttgaaacaaccaatcaaaataaagat
gaagtgaatggaaaagaacaaaatcaagcaaatcttcagtaacaaatacacaaattacc
aaaaacgagaaaagacgaagacacaaaaacactaaagaaagataaagacgagaaaagaatct
25 aaagacacaaaaacaccaaagaaagacaaaagaaagacataaaaaactccgaagaaa
aagagaaagagaaaaaacagtaataccaaaaagcggcaagacgagaaaagacacaaaa
ataactaagaaagacaaaagaaagacgaaattacaacaacttccaagaaagataataacaat
gatgtacaagataaattaccggaacaggtaaaacaaacgatattcaaaatcctgcttta
ataatgttacttgctggtttaggtttatttaggattatttagaaataaaataagagaatag
30

30

Sequence 2008

35 MKDNKPNNKLIQTYLSKKTLYGTASALTALYLFNSNVTYADENTANQNQGTSPKTS
QTAPTNNNTENTDATAITTDQNNNDEEYDASYELPILYVTWVLDQGNIIKDAVEDAKTP
ASERQPVKIPGYQHYRTSVSDGITKFIYRKISTAQSPIVENNQDNNTNKVETTNQNKD
EVNGKEQNQANTSVTNTQITKNEKDEDTKTLKKDKDEKESKDTKTPKKDKKKDIKTPKK
DREEKKPVIKSGKDEKDKITKDKDEDEITTSKKDNNNDVQDKLPETGKTNDIQNPAL
IMLLAGLGLLGLFRNKIRE*

40 Sequence 2009

Contig_0693_pos_3675_1084,

putative peptide of unknown function

gtgaaattaccttatggtgtgcaacaagacgctcatgaagtagaagatgcacttgagttt
attaatcgtgtaattacacctttatcaccgatttcaacatttgctgcccgtaatccgtgg
45 gaggggctagaagatgcttcggttgatcaagtggcagcttggttaaaaagtgtgagggat
gttgacatttatcctaatacgctctactattcacagagcgattagtaataaagaaatagat
ttaaagtatattgaagaacggttggtgaaaatcgctgcgcattataataataggtcacta
tctgacagtgatatcaacacatatattcaagagcgaaaaatttaaaaacgattgaagaa
ggttactttaatacaaaaagataacgagaaaactggaaaaatgggtacaaactaattttaag
50 gattataagaaaaaagaagatgtgatagcgcaaagtgctagtgtttcacaaggaaggt
acacgacttattgatattttaaatgctcatatgattaagtggtctaaatttatatgttgat
gactttcaatcaagttggactatgccaaaaagagaaaaaggattctatcatgcctggcaa
cgtttagttaaacatgatccattattcacaaaaaacaacgacttacttttagcacatttg
ccaatcaagcaaccgaagcaatagagtagcgcttcaagaattaggagtaaaagaagaa
55 catcgacaatcatatattgagagtcattttattatctttaccagggttgggcaggaatcatg
tatcatcggtcacagacacaaagtaatgatgcgtacttattaacagactatgttgcgatt
cgtctatcaattgagatggtacttttaaatgaccaccatacaacattattaaaaaaatct
atatatcttcaaaaaaagttagagcaaatacgttatttgctatttaacatacaaatgaat
gttgagcagtggttaaatctatcatctaaaaagcaacaagcatatcattgaattggggaca